

REC'D 14 OCT 2003

WIPO PCT

PA 1057875

# THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 28, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: 60/432,259

FILING DATE: December 11, 2002

## PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the  
COMMISSIONER OF PATENTS AND TRADEMARKS



T. LAWRENCE  
Certifying Officer

# Provisional Application For Patent Cover Sheet

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 C.F.R. § 1.53(c).

Express Mail Label No.

## INVENTOR(S)

Given Name (first and middle (if any))	Family Name or Surname	Residence (City and either State or Foreign Country)
Lidia Wolfgang A. Philippe	IVANOVA RENNER SAUDAN	Schlieren, Switzerland Kilchberg, Switzerland Pfungen, Switzerland

PTO  
65234/09  
12/11/02  
JC996 U.S. 9660

☐ Additional inventors are being named on the \_\_\_\_\_ separately numbered sheets attached hereto.

## TITLE OF THE INVENTION (500 Characters Maximum)

Gene Expression System

## CORRESPONDENCE ADDRESS

Direct all correspondence to:

☒ Customer number 26111  
Type Customer Number here

Place Customer Number

Bar Code Label here



26111

PATENT TRADEMARK OFFICE

OR

☐ Firm or Individual Name

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

Address

Address

City

State

ZIP

Country

USA

Telephone

Fax

## ENCLOSED APPLICATION PARTS (check all that apply)

- ☒ Specification Number of pages 79 ☐ CD(s), Number \_\_\_\_\_
- ☒ Drawing(s) Number of sheets: 9 ☒ Other (specify) Sequence Listing - 33 pages
- ☒ Application Data Sheet. See 37 CFR 1.76

## METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (check one)

- ☐ Applicant claims small entity status. See 37 CFR 1.27.
- ☐ A check or money order is enclosed to cover the filing fees
- ☒ The Commissioner is hereby authorized to charge filing fees or credit any overpayments to Deposit Account Number: 19-0036.
- ☒ Payment by credit card. Form PTO-2038 is attached.

Filing Fee  
Amount (\$) 160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

- ☒ No
- ☐ Yes, the name of the U.S. Government agency and the Government contract number are:

Respectfully submitted,  
Signature: Frank R. Cottingham  
Typed or Printed Name: Frank R. Cottingham  
Telephone: 202-371-2600

Date: 12/11/02  
Registration No. 50.437 (if appropriate)  
Docket Number: 1700.0370001/JAG/FRC

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

SKGE DCI-83095.1

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid MB control number

Approved for use through 10/31/2002 OMB 0651-0032  
Patent and Trademark Office U.S. DEPARTMENT OF COMMERCE  
PTO/SB/17 (10-02)

# FEE TRANSMITTAL

f r FY 2003

Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT (\$) 160.00

## Complete if Known

Application Number	To Be Assigned
Filing Date	Herewith
First Named Inventor	Lidia IVANOVA
Examiner Name	N/A
Group Art Unit	N/A
Attorney Docket No.	1700.0370001/JAG/FRC

## METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit card ☐ Money Order ☒ Other\* ☐ None

☐ Deposit Account

Deposit Account Number 19-0036

Deposit Account Name Sterne, Kessler, Goldstein & Fox P.L.L.C.

The Commissioner is authorized to: (check all that apply)

☒ Charge fee(s) indicated below ☒ Credit any over payments

☒ Charge any additional fee(s) during the pendency of this application

☐ Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account

\*\* Charge any deficiencies or credit any overpayments in the fees or fee calculations of Parts 1, 2 and 3 below to Deposit Account No 19-0036.

## FEE CALCULATION

### 1. BASIC FILING FEE

Large Entity Small Entity

Fee Code	Fee (\$)	Fee Code	Fee (\$)	Fee Description	Fee Paid
1001	740	2001	370	Utility filing fee	
1002	330	2002	165	Design filing fee	
1003	510	2003	255	Plant filing fee	
1004	740	2004	370	Reissue filing fee	
1105	160	2005	80	Provisional filing fee	160.00

SUBTOTAL (1) (\$) 160.00

### 2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

Total Claims	Extra Claims	Fee from below	Fee Paid
Independent	- 20** =	x	=
Multiple Dependent	- 3** =	x	=

Large Entity Small Entity

Fee Code	Fee (\$)	Fee Code	Fee (\$)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	**Reissue independent claims over original patent
1205	18	2205	9	**Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) (\$) 0.00

\*\* or number previously paid, if greater; For Reissue, see above

## FEE CALCULATION (continued)

### 3. ADDITIONAL FEES

Large Entity Small Entity

Fee Code	Fee (\$)	Fee Code	Fee (\$)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1502	50	2052	25	Surcharge-late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for ex parte reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	400	2252	200	Extension for reply within second month	
1253	920	2253	460	Extension for reply within third month	
1254	1,440	2254	720	Extension for reply within fourth month	
1255	1,960	2255	980	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing a brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1,280	2453	640	Petition to revive - unintentional	
1501	1,280	2501	640	Utility issue fee (or reissue)	
1502	460	2502	230	Design issue fee	
1503	620	2503	310	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	740	2809	370	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	740	2810	370	For each additional invention to be examined (37 CFR 1.129(b))	
1801	740	2801	370	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify)

\* Reduced by Basic Filing Fee Paid

SUBTOTAL (3) (\$) 0.00

## SUBMITTED BY

Name (Print/Type)	Frank R Cottingham	Registration No (Attorney/Agent)	50,437	Complete (if applicable)
Signature	<i>Frank R. Cottingham</i>	Telephone	202-371-2600	Date
				12/11/02

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

This collection of information is required by 37 CFR 1.17 and 1.27. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, Washington, DC 20231.

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2



**Sterne Kessler  
Goldstein Fox**  
ATTORNEYS AT LAW

Robert Greene Sterne  
Edward J. Kessler  
Jorge A. Goldstein  
David K.S. Cornwell  
Robert W. Esmond  
Tracy-Gene G. Durkin  
Michele A. Cimbala  
Michael B. Ray  
Robert E. Sokohl  
Eric K. Steffe  
Michael Q. Lee  
Steven R. Ludwig  
John M. Covert  
Linda E. Alcorn  
Robert C. Millonig  
Lawrence B. Bugalsky  
Donald J. Featherstone  
Michael V. Messinger

Judith U. Kim  
Timothy J. Shea, Jr.  
Patrick E. Garrett  
Jeffrey T. Helvey  
Heshi L. Kraus  
Crystal D. Sayles  
Edward W. Yee  
Albert L. Ferro  
Donald R. Banowit  
Peter A. Jackman  
Molly A. McCall  
Teresa U. Medler  
Jeffrey S. Weaver  
Kendrick P. Patterson  
Vincent L. Capuano  
Albert J. Passulo II  
Eldora Ellison Floyd  
Thomas C. Fiala

Brian J. Del Buono  
Virgil Lee Beaton  
Kimberly N. Reddick  
Theodore A. Wood  
Elizabeth J. Haanes  
Bruce E. Chalker  
Joseph S. Ostroff  
Frank R. Cottingham  
Christine M. Lhulier  
Rae Lynn Prengaman  
Jane Shershenovich  
Lawrence J. Carroll  
George S. Bardmesser  
Daniel A. Klein  
Rodney G. Maza  
Jason D. Eisenberg  
Michael A. Specht  
Andrea J. Kamage

**Registered Patent Agents\***  
Karen R. Markowicz  
Nancy J. Lenth  
Ann E. Summerfield  
Helene C. Carlson  
Gaby L. Longworth  
Matthew J. Dowd  
Aaron L. Schwartz  
Angelique G. Uy  
Boris A. Matvenko  
Mary B. Tung  
Katrina Y. Pei  
Bryan L. Skelton  
Robert A. Schwartzman  
John J. Figueroa  
Timothy A. Doyle  
Jennifer R. Mahalingappa  
Jeff Lungren

Teresa Colella  
Jon Wright  
Victoria Rutherford  
W. Brian Edge

**Senior Counsel**  
Samuel L. Fox  
Kenneth C. Bass  
Lisa A. Dummer

\*Admitted only in Maryland  
\*Admitted only in Virginia  
\*Admitted only in Texas  
\*Practice limited to  
Federal Agencies

PTO  
JC996 U.S. PTO  
60/432259

December 11, 2002

**WRITER'S DIRECT NUMBER:**  
(202) 371-2615

**INTERNET ADDRESS:**  
FRANKC@SKGF.COM

Commissioner for Patents  
Washington, D.C. 20231

**Box Provisional Application**

Re: U.S. Provisional Patent Application  
Appl. No. To Be Assigned; Filed: Herewith  
For: **Gene Expression System**  
Inventors: Ivanova *et al.*  
Our Ref: 1700.0370001/JAG/FRC

Sir:

The following documents are being submitted under 37 C.F.R. § 1.53(c) herewith for appropriate action by the U.S. Patent and Trademark Office:

1. PTO Fee Transmittal (Form PTO/SB/17);
2. U.S. Provisional Patent Application entitled:

**Gene Expression System**

and naming as inventors:

Lidia IVANOVA  
Wolfgang A. RENNER  
Philippe SAUDAN

the application consisting of:

- a. A Provisional Application for Patent Cover Sheet;
- b. An Application Data Sheet (37 C.F.R. § 1.76);
- c. A specification containing 79 total pages:

Commissioner for Patents

December 11, 2002

Page 2

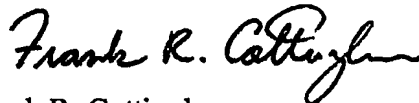
- (i) 62 pages of description prior to any claims;
- (ii) 16 pages of claims (109 claims); and
- (iii) a one page abstract;
- d. 9 sheets of drawings: (Figures 1, 2, 3, 4, 5, 6, 7, 8 and 9 ); and
- e. 33 pages of a paper copy of a sequence listing after the abstract;
- 3. Authorization to Treat a Reply As Incorporating An Extension of Time Under 37 C.F.R. § 1.136(a)(3);
- 4. Form PTO-2038 Credit Card Payment From in the amount of \$ 160.00 to cover the filing fee; and
- 5. Two (2) return postcards.

It is respectfully requested that, of the two attached postcards, one be stamped with the filing date of these documents and returned to our courier, and the other, prepaid postcard, be stamped with the filing date and returned as soon as possible.

The U.S. Patent and Trademark Office is hereby authorized to charge any fee deficiency, or credit any overpayment, to our Deposit Account No. 19-0036.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.



Frank R. Cottingham  
Attorney for Applicants  
Registration No. 50,437

Enclosures

::ODMA\MIHODMASKGF\_DC1;83026;1

INVENTOR INFORMATION

Inventor One Given Name:: Lidia  
Family Name:: Ivanova  
Postal Address Line One:: Schulstrasse 31  
City:: CH-8952 Schlieren  
Country:: Switzerland  
City of Residence:: CH-8952 Schlieren  
Country of Residence:: Switzerland  
Citizenship Country:: U.S.A.  
Inventor Two Given Name:: Wolfgang A  
Family Name:: Renner  
Postal Address Line One:: Grenzsteig 9  
City:: CH-8802 Kilchberg  
Country:: Switzerland  
City of Residence:: CH-8802 Kilchberg  
Country of Residence:: Switzerland  
Citizenship Country:: Switzerland  
Inventor Three Given Name:: Philippe  
Family Name:: Saudan  
Postal Address Line One:: Durrenrainstrasse 20  
City:: CH-8422 Pfungen  
Country:: Switzerland  
City of Residence:: CH-8422 Pfungen  
Country of Residence:: Switzerland  
Citizenship Country:: Switzerland

CORRESPONDENCE INFORMATION

Correspondence Customer Number:: 26111

APPLICATION INFORMATION

Title Line One:: Gene Expression System  
Total Drawing Sheets:: 9  
Formal Drawings?: No  
Application Type:: Provisional  
Docket Number:: 1700.0370001  
Secrecy Order in Parent Appl.?: No

REPRESENTATIVE INFORMATION

Representative Customer Number:: 26111  
Source:: PrintEFS Version 1.0.1

## GENE EXPRESSION SYSTEM

### BACKGROUND OF THE INVENTION

#### Field of the Invention

[0001] The present invention relates to compositions and methods for the production of polypeptides and/or untranslated RNA molecules in host cells. The invention provides nucleic acid molecules, expression systems, host cells, methods and kits that are useful for the production of polypeptides and/or untranslated RNA molecules. In a preferred embodiment, the invention allows regulated and/or inducible expression of a nucleotide sequence of interest in a host cell.

#### Related Art

[0002] Process development for biopharmaceuticals is influenced by product quality and economy of the manufacturing process. The economic production of recombinant proteins in mammalian cells is dependent on the selection of the producing cell lines. The classical approach is the use of stable expression systems. These systems are based on chromosomal integration of an expression plasmid into the genome of the host cell.

[0003] An alternative to stable expression systems is transient gene expression. For example, transient gene expression in, e.g., mammalian cells at reactor scale is becoming increasingly important for the rapid production of recombinant proteins. Large scale applications of transient expression systems for recombinant protein production have been reported using COS, HEK 293 and BHK21 cells (Blasey et al., 1996, *Cytotechnology*, 18, 138-192; Jordan et al., 1998, *Cytotechnology*, 26, 39-47; Wurm and Bernard, 1999, *Curr Opin Biotechnol* 10, 156-159; Meissner et al., 2001, *Biotechnol and Bioeng*, 75, 2, 197-203).

[0004] Transient expression can be performed by infection (using, e.g., viral vectors) or by transfection of the appropriate vector. Transfection can be

**[0008]** The double strand DNA genome of Epstein-Barr virus EBV (165 kb) is maintained in the latently infected cells as a large circular episome. Amplification of this episome is mediated by a cis-acting sequence such as the



origin of replication (OriP), and one of the viral proteins, the Epstein-Barr viral nuclear antigen 1 (EBNA1). In a minimal system the presence of both OriP (1.8 kb region) and EBNA-1 is sufficient for replication and stabilization of recombinant plasmids. The exact function of EBNA-1 is still not clear, although the protein has a DNA binding domain. Generally EBNA-1 is involved in plasmid maintenance after completion of the synthesis and in plasmid segregation by dividing cells during proliferation. The OriP region is recognized also directly by the cellular DNA synthesizing machinery and is replicated during the cell cycle.

- [0009] Other gene expression systems are based on alphaviruses (Lundstrom, K., *Curr. Opin. Biotechnol.* 8:578-582 (1997)). Several members of the alphavirus family, e.g., Sindbis virus (Xiong, C. *et al.*, *Science* 243:1188-1191 (1989); Schlesinger, S., *Trends Biotechnol.* 11:18-22 (1993)), SFV (Liljeström, P. & Garoff, H., *Bio/Technology* 9:1356-1361 (1991)) and others (Davis, N.L. *et al.*, *Virology* 171:189-204 (1989)) have received considerable attention for the use as virus-based expression vectors for a variety of different proteins (Lundstrom, K., *Curr. Opin. Biotechnol.* 8:578-582 (1997); Liljeström, P., *Curr. Opin. Biotechnol.* 5:495-500 (1994)).
- [0010] Alphaviruses are positive stranded RNA viruses which replicate their genomic RNA entirely in the cytoplasm of the infected cell and without a DNA intermediate (Strauss, J. and Strauss, E., *Microbiol. Rev.* 58:491-562 (1994)). The concept that alphaviruses can be developed as expression vectors was first established nearly ten years ago (Xiong, C. *et al.*, *Science* 243:1188-1191 (1989)). Since then, several improvements have made the use of these RNA replicons as expression vectors more practical (Lundstrom, K., *Curr. Opin. Biotechnol.* 8:578-582 (1997)).
- [0011] DNA vectors have been developed for both Sindbis virus (Herweijer, H. *et al.*, *Hum. Gene Ther.* 6:1495-1501 (1995); Dubensky, T.W. *et al.*, *J. Virol.* 70:508-519 (1996)) and SFV (Berglund, P. *et al.*, *Trends Biotechnol.* 14:130-134 (1996)). Eukaryotic promoters are introduced in these vectors upstream from the alphavirus replicase gene (consisting of the four non-structural protein genes (nsP1-4)), which are translated as one or two

polyproteins which are then proteolytically cleaved. DNA is transcribed into RNA from the recombinant eukaryotic promoter in the nucleus and transported to the cytoplasm, where the replicase catalyzes the replication of the alphavirus RNA molecule as during normal replication of the alphavirus RNA molecule (Strauss, J. and Strauss, E., *Microbiol. Rev.* 58:491-562 (1994)). Only transient expression of heterologous sequences has been possible until recently due to the cytopathogenicity of the alphavirus replicase (Lundstrom, K., *Curr. Opin. Biotechnol.* 8:578-582 (1997)).

[0012] Among the features which make Alphavirus vectors attractive as transient expression systems are the easy construction of plasmid clones, a rapid gene expression (within 24h), a high level of expression, a wide host range of the virus (many insect, avian and mammalian cells), production of high-titered stocks of infectious particles and low risk of cell transformation due to the RNA nature of the genome and its replication entirely in the host cell cytoplasm without going through DNA intermediates. (Schlesinger S., 2000, *Exp. Opin. Biol. Ther.*, 1, 2, 177-191; Schlesinger S., *Adv. Virus res.*, v55, 565-577; Garoff, H. and K.J. Li, 1998, *Curr. Opin. Biotechnol.*, 9, 464-469; Lundstrom, K., 1997, *Curr. Opin. Biotechnol.*, 8, 578-583; Xiong et al., 1989, *Science*, 243, 1188-1191). Blasey et al. compared the Semliki Forest Virus (SFV) expression system with the Epstein-Barr Virus (EBV) system (pCEP4 and HEK293EBNA cells) for expression of 5-HT3 receptors and found that both systems allow the production of functional recombinant protein within days to weeks at very high comparable levels (in: *Animal Cell Technology*, 1999, pp331-337).

[0013] One of the variants of alphavirus vectors is the replicon, which is capable of self-replication in the host cytoplasm after introduction into the cultured cells by transfection. In this replicon, viral structural protein genes have been replaced by a gene of interest (GOI). Such a replicon can be introduced into cells as RNA or cDNA if placed under the control of an eukaryotic promoter or a promoter, which is recognized by the host cell machinery. (Agapov et al., *Proc. Natl. Acad. Sci. USA* 95:12989-12994 (1998)).

- [0014] RNA viruses have been engineered into chimeric viruses producing heterologous proteins. Such RNA virus vectors have been developed from Alphaviruses (with positive strand genome), Influenza and VSV (with negative strand genome). One of the major advantages of these viruses is their RNA-only replication strategy performed entirely in the cell cytoplasm and eliminating the risk of genome integration.
- [0015] Adeno-, Adeno-associated and Retrovirus vectors can be used as gene therapy vectors and/or as expression vectors, especially for large-scale protein production. Herpesvirus derivative vectors containing a replication origin have been used for bioreactor protein production after transient transfection of a suitable cell line carrying a replication initiation gene. (Durocher *et al.*, *Nucleic Acids Res.* 30(2):E9 (2002); Meissner *et al.*, *Biotechnol. Bioeng.* 75(2):197-203 (2001)). In these experiments, 293E cells (293 human embryo kidney cells expressing Epstein-Barr virus nuclear antigene - EBNA-1) were used which provided a replication initiation factor in trans to a plasmid carrying EBV origin of replication - OriP. Reporter genes (encoding secreted, intracellular or transmembrane proteins) were constitutively expressed over a period of several days.
- [0016] Hybrid vectors have also been reported, mainly as gene delivery systems or for use in virus biology research, such as combinations between EBV episomes and Herpes-, Adeno- (DNA) or Retro- (RNA-DNA) viruses (Sena-Esteves M. *et al.*, 1999, *Virol. Dec*;73(12):10426-39; Tan BT. *et al.*, 1999, *J Virol. Sep*;73(9):7582-9; Grignani F. *et al.*, 1998, *Cancer Res Jan* 1;58(1):14-9; Wang S. and Vos JM., 1996, *J Virol. Dec*;70(12):8422-30). EBV episomes are nuclear plasmids that are stably maintained through multiple cell divisions in primate and canine cells (J. L. Yates, N. Warren, and B. Sugden, 1985, *Nature* 313:812-815).
- [0017] The ability to precisely control the expression of genes introduced into animal or human cells, or in whole organisms, enables significant progress in many areas of biology and medicine. For instance, methods that allow the intentional manipulation of gene expression will facilitate the analysis of genes whose expression cannot be tolerated constitutively.

[0018] To be of broad benefit, gene regulation techniques preferably allow for rapid, robust, precise and reversible induction of gene activity. As reviewed in Saez, E. *et al.*, (*Curr. Opin. Biotechnol.* 8:608-616 (1997)), an ideal system would fulfill the following requirements:

[0019] 1. Specificity -- The system should be indifferent to endogenous factors and activated only by exogenous stimuli.

[0020] 2. Non-interference -- The components of the system should not affect unintended cellular pathways.

[0021] 3. Inducibility -- In the inactive state, the basal activity of the system should be minimal, while in the active state high levels of gene expression should be rapidly inducible.

[0022] 4. Bioavailability of the inducer -- Inducing stimuli should rapidly penetrate to the site of interest.

[0023] 5. Reversibility -- Inducing stimuli should clear swiftly to allow the system to rapidly return to the inactive state.

[0024] One common system currently in use for the regulation of gene expression is the tetracycline-based system (Gossen and Bujard, *Proc. Natl. Acad. Sci. USA* 89:5547 (1992)). This system is based on the continuous expression of a fusion protein with the tetracycline repressor protein (tetR) being converted into an activator by fusion to the transcriptional activation domain of the VP16 protein. The tTA system is useful for, *e.g.*, inducible gene expression and has been successfully used for the production of a number of proteins (Wimmel *et al.*, *Oncogene* 9:995 (1994); Fröh *et al.*, *EMBO J.* 13:3236 (1994); Yu *et al.*, *J. Virol.* 70:4530 (1996)). However, there are certain disadvantages associated with the tTA system. (Schocket *et al.*, *Proc. Natl. Acad. Sci. USA* 92:6522 (1995); Howe *et al.*, *J. Biol. Chem.* 23:14168 (1995); Schocket and Schatz, *Proc. Natl. Acad. Sci. USA* 93:5173 (1996); Bohl *et al.*, *Nat. Med.* 3:299 (1997); Furth *et al.*, *Proc. Natl. Acad. Sci. USA* 91:9302 (1994); Hennighausen *et al.*, *J. Cell. Biochem.* 59:463 (1995); Kistner *et al.*, *Proc. Natl. Acad. Sci. USA* 93:10933 (1996); Hoffmann *et al.*, *Nucleic Acids Res.* 25:1078 (1997)).

-7-

[0025] Inducible expression systems based on viral expression vectors (*e.g.*, alphaviral vectors) are described in WO 99/50432 (incorporated herein by reference in its entirety). An exemplary alphaviral vector is pCytTs (also known as CytTs; *cf.* WO 99/50432). Using these systems, it is possible to precisely induce and regulate gene expression. This precise regulation results from the use of a temperature-sensitive RNA-dependent RNA polymerase (*i.e.*, a replicase), which only replicates RNA molecules, to form new RNA molecules, at permissive temperatures (Boorsma et al. *Nature Biotech.* 18:429-432 (2001)). Due to the extremely tight regulation of these systems it is possible to obtain regulated gene expression even under conditions of transient transfection allowing the production of toxic proteins under these conditions.

[0026] Another example of an alphaviral expression vector is described in Lundstrom *et al.*, *FEBS Lett* 504:99-103 (2001). Lundstrom describe a rapidly generated high-titer Semliki Forest virus (SFV) vectors capable of infecting numerous mammalian cell lines and primary cell cultures, resulting in high levels of transgene expression. SFV-based expression of transmembrane receptors has been characterized by specific ligand-binding activity and functional responses. Adaptation of the SFV technology for mammalian suspension cultures has allowed the production of hundreds of milligrams of recombinant receptor for purification and structural studies. The same SFV stock solutions used for the infection of mammalian cells in culture have also been successfully applied for efficient transgene expression in organotypic hippocampal slices, as well as in vivo in rodent brain.

[0027] A need exists in the art, however, for compositions and methods that allow the production of polypeptides and untranslated RNA molecules.

#### BRIEF SUMMARY OF THE INVENTION

[0028] The present invention provides compositions and methods that allow the production of polypeptides and/or untranslated RNA molecules. More specifically, the invention provides nucleic acid molecules, expression systems, recombinant host cells that permit the production of polypeptides

and/or untranslated RNA molecules. Also provided are methods for making the nucleic acid molecules, expression systems and host cells of the invention. The invention further provides methods for producing polypeptides and/or untranslated RNA molecules. The invention also provides kits that comprise the nucleic acid molecules, expression systems and/or recombinant host cells of the invention.

[0029] In certain embodiments, the invention provides compositions and methods for regulated, expression of polypeptides or untranslated RNA molecules in recombinant host cells, preferably by way of transfection. The present invention also provides nucleic acid molecules and methods which allow, *e.g.*, rapid and high level production of specific RNA molecules produced in, *e.g.*, transiently transfected, recombinant host cells.

[0030] According to certain aspects of the invention, novel, preferably inducible, potent and rapid expression systems are provided comprising one or more nucleic acid molecules. The nucleic acid molecules of the expression systems of the invention may be maintained at, or accumulated to, multiple copies in the nuclei of recombinant host cells, preferably of transiently transfected mammalian cells. In certain embodiments, RNA self replication may lead to the accumulation of a high number of RNA molecules in the cytoplasm of the transfected cells and subsequent translation of the polypeptide of interest. The nucleic acid molecules and expression systems of the invention are particularly useful for transient transfection.

[0031] The present invention further provides nucleic acid molecules and methods which allow inducible rapid and high level production of specific RNA molecules produced in, preferably transiently transfected, recombinant host cells. In certain embodiments, RNA self-replication is preferably inducible and triggered by a temperature shift. Thus, the high production levels, in certain embodiments, are achieved by the use of a temperature-sensitive RNA-dependent RNA polymerase (*i.e.*, a replicase) which only replicates RNA molecules to form new RNA molecules at permissive temperatures. Moreover, if desired, the amount of RNA or polypeptide can be tightly regulated by the use of a temperature-sensitive RNA-dependent RNA

polymerase. This facilitates, in particular, production of growth-inhibitory or toxic polypeptides or RNA-species.

[0032] In a general aspect, the present invention provides for a nucleic acid molecule comprising (A) a first polynucleotide element which encodes an RNA molecule, said RNA molecule comprising (a) at least one *cis*-acting sequence element; (b) a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and (c) at least one second nucleotide sequence selected from the group consisting of (i) a second open reading frame encoding a polypeptide; (ii) a nucleotide sequence complementary to all or part of the second open reading frame of (i); and (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof; wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase; (B) a second DNA element comprising an origin of replication; and (C) a third DNA element encoding a replication initiation factor capable of recognizing the origin of replication. Thus, the replication initiation factor preferably recognizes the origin of replication leading to replication of the second polynucleotide and second DNA element, respectively, comprising the origin of replication.

[0033] In certain embodiments, the RNA-dependent RNA polymerase is, *e.g.*, a non-cytopathic RNA-dependent RNA polymerase. In other embodiments, the RNA-dependent RNA polymerase is, *e.g.*, a temperature-sensitive RNA-dependent RNA polymerase. In yet other embodiments, the RNA-dependent RNA polymerase is, *e.g.*, a non-cytopathic, temperature-sensitive RNA-dependent RNA polymerase.

[0034] The nucleic acid molecules of the invention, in certain embodiments, comprise a 5' promoter which is capable of initiating transcription *in vivo*, 5' and/ or 3' sequences enabling replication of the RNA molecule (*cis*-acting sequence elements), and a subgenomic promoter 5' to the gene of interest, as well as a sequence of interest which is translatable only after one or more RNA-dependent RNA replication events. These RNA-dependent RNA





**production of toxic polypeptides if the RNA virus replication is controllable or inducible.**

**[0038]** In a specific embodiment, the invention includes the combination of a Herpesvirus mini-replicon unit, i.e. the cis-acting replication origin OriP (Origin of replication P) and the cis- or trans-acting gene product – EBNA-1 (Epstein-Barr virus nuclear antigen 1, a replication triggering factor) with the tightly regulated temperature inducible alphaviral expression system pCytTS, as disclosed in WO 99/504332. Due to the presence of OriP and EBNA-1 the introduced novel vector is maintained episomally (extrachromosomally) in the form of several DNA copies in the cell nuclei. These multiple DNA copies may be transcribed from a promoter, typically a CMV or RSV promoter, into CytTS RNA-replicons, but they remain inactive unless cells are being shifted to a certain temperature. Only after temperature induction replicon replication takes place, followed by RNA accumulation and translation in the cell cytoplasm and production of the polypeptide of interest.

[0039] In a further general aspect, the present invention provides a vector system comprising one or more nucleic acid molecules, wherein said one or more nucleic acid molecules comprise (A) a first polynucleotide which encodes an RNA molecule, said RNA molecule comprising (a) at least one *cis*-acting sequence element; (b) a first nucleotide sequence comprising a first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and (c) at least one second nucleotide sequence selected from the group consisting of (i) a second open reading frame encoding a polypeptide; (ii) a sequence complementary to all or part of the second open reading frame of (i); and (iii) a sequence encoding an untranslated RNA molecule, or complement thereof; wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said non-cytopathic RNA-dependent RNA polymerase; (B) a second polynucleotide comprising an origin of replication; and (C) a third polynucleotide encoding a replication initiation factor capable of recognizing the origin of replication. Thus, the replication initiation factor recognizes the

origin of replication leading to replication of the second polynucleotide and second DNA element, respectively, comprising the origin of replication.

[0040] The invention further provides single- and multiple-vector systems for producing a polypeptide or untranslated RNA molecule. In a single-vector system, the first, second and third polynucleotide are present on the same nucleic acid molecule. In a multiple-vector system, the first, second and third polynucleotide elements are present on one or more separate nucleic acid molecules. In such a multiple vector system, at least one, preferably some or more preferably all vectors contain an origin of replication. Therefore, in preferred embodiments of the invention, the replication initiation factor and the origin of replication recognizing the replication initiation factor may be on the same nucleic acid molecule and plasmid, respectively, as it is the RNA-replicase or on a separate nucleic acid molecule and plasmid, respectively.

[0041] When sequences encoding the first and second open reading frame are present either on the same nucleic acid molecule or in the same vector (i.e., in a single-vector system), a region will preferably be present 5' to the second open reading frame which inhibits translation of this open reading frame.

[0042] If a temperature sensitive RNA-dependent RNA polymerase (*e.g.*, a temperature-sensitive replicase) is included within the compositions and methods of the invention, the temperature-sensitive replicase may be "cold" or "hot" sensitive and thus will only efficiently catalyze RNA-dependent RNA replication at temperatures either above or below the restrictive temperature. In one preferred embodiment, the compositions and methods of the invention comprise, or involve the use of, an RNA-dependent RNA polymerase that has replicase activity at temperatures below 34°C which is at least five fold greater than the replicase activity exhibited at 34°C or above. Preferably, the replicase activity at temperatures below 34°C is at least ten fold greater than the replicase activity exhibited at 34°C or above. In certain embodiments, the replicase activity at temperatures below 34°C is at least one hundred fold greater than the replicase activity exhibited at 34°C or above. Temperature-sensitive RNA-dependent RNA polymerases included within the present

invention may, in certain embodiments, exhibit no (or undetectable) replicase activity at 34°C or above.

- [0043]** In another aspect, the nucleic acid molecules and expression systems of the invention encode one or more cytokine, lymphokine, tumor necrosis factor, interferon, toxic protein, prodrug converting enzyme, or other polypeptide.
- [0044]** In yet another aspect, the nucleic acid molecules and expression systems of the invention encode an untranslated RNA molecule, such as an antisense RNA molecule, tRNA molecule, rRNA molecule, or ribozyme.
- [0045]** The invention also provides methods for making recombinant host cells comprising introducing nucleic acid molecules of the invention into host cells. Further provided are recombinant host cells produced by the introduction of nucleic acid molecules of the invention. In one embodiment, one, some or all of these recombinant host cells contain one or more nucleic acid molecules that comprise an RNA-dependent RNA polymerase and/or the replication initiation factor. The replication initiation factor, in certain embodiments, may be stably integrated into the genome of the one, some or all host cells.
- [0046]** The invention further provides isolated nucleic acid molecules comprising polynucleotides which comprise the nucleotide sequence of SEQ ID NO:21 (pCytTs-OrIP) and SEQ ID NO:22 (pCytTs-OPE).
- [0047]** The present invention also provides methods for producing polypeptides and untranslated RNA molecules in recombinant host cells, said methods comprising introducing a nucleic acid molecule of the invention into a host cell to produce a recombinant host cell, and culturing the recombinant host cell under conditions suitable for expression of said polypeptide or untranslated RNA molecule. The methods of the invention may further comprise recovering said polypeptide or untranslated RNA molecule.
- [0048]** Methods are also provided for the expression of heterologous polypeptides, preferably the regulated expression of heterologous

polypeptides, including cytokines, lymphokines, tumor necrosis factors, interferons, toxic polypeptides, and prodrug converting enzymes.

[0049] Further provided are polypeptides and untranslated RNA molecules produced by the methods of the invention.

[0050] The invention also provides methods for regulating the expression of a polypeptide or an untranslated RNA molecule, said method comprising introducing nucleic acid molecules of the invention into a host cell to produce a recombinant host cell and growing said recombinant host cell under suitable culture conditions. In certain embodiments, nucleic acid molecules encoding a temperature-sensitive RNA-dependent RNA-polymerase are used. When a temperature-sensitive RNA-dependent RNA-polymerase is used, the methods according to this aspect of the invention may further comprise changing the temperature of the recombinant host cell culture from either (i) a permissive temperature to a restrictive temperature, or (ii) a restrictive temperature to a permissive temperature.

[0051] In certain embodiments, the methods of the invention involve introducing the nucleic acid molecules of the invention into prokaryotic or eukaryotic host cells to produce a recombinant host cell, and then culturing said recombinant host cell under suitable culture conditions. The recombinant host cells may be cultured, *e.g.*, in a serum-free or protein-free medium.

[0052] The present invention also provides pharmaceutical compositions comprising nucleic acid molecules of the invention and a pharmaceutically acceptable carrier.

[0053] Also included within the invention are kits comprising the nucleic acid molecules of the invention. Kits of the invention may additionally or alternatively comprise one or more expression systems of the invention and one or more recombinant host cells of the invention.

#### BRIEF DESCRIPTION OF THE DRAWING

[0054] Fig. 1 is a schematic representation of the pCytTs vector. The pCytTs vector contains an ampicillin resistance marker for selection in bacterial cells

and a ColE1 sequence which directs high copy number bacterial amplification. The pCytTs vector was prepared as described in Example 1. MCS: multiple cloning site; RSV: Rous Sarcoma Virus promoter, amp: ampicillin resistance marker; ColE1: bacterial origin of replication; nsP1-4: genes coding for the viral non-structural proteins; Pro728Ser: nsP2 mutation that renders the replicase non-cytopathic; Gly153Glu: nsP4 mutation that renders the replicase temperature-sensitive; Psg: subgenomic promoter.

[0055] Fig. 2 is a schematic representation of vector pCytTS2.1. The pCytTS2.1 vector contains an ampicillin resistance marker for selection in bacterial cells and a ColE1 sequence which directs high copy number bacterial amplification. The pCytTS2.1 vector was prepared as described in Example 1. MCS: multiple cloning site; CMV: Cytomegalovirus promoter, puro: puromycin resistance marker; amp: ampicillin resistance marker; ColE1: bacterial origin of replication; nsP1-4: genes coding for the viral non-structural proteins; Pro728Ser: nsP2 mutation that renders the replicase non-cytopathic; Gly153Glu: nsP4 mutation that renders the replicase temperature-sensitive; Psg: subgenomic promoter.

[0056] Fig. 3 is a schematic representation of vector pCytTs-EGFP-OPE described in Example 2. The EGFP expression cassette can be replaced by any gene of interest (e.g. SEAP, EPO, IFN). For figure legend see Fig. 2. GFP: green fluorescence protein.

[0057] Fig. 4 demonstrates CytTs-EGFP-OriPd derived expression in 293-EBNA cells. Selected cell population of 293-EBNA cells, which had been transfected with CytTs-EGFP-OriPd and selected with puromycin were either kept at 37°C or shifted to 29°C for 24 hours in order to induce the Sindbis replicon. Cells were trypsinised and analyzed by flow cytometry. Untransfected and CytTs-EGFP-OriPd transfected 293-EBNA cells either cultivated at 37°C or at 29°C, as indicated above the respective panels were analyzed. GFP expression is shown on the X-axis (FL1-H) the Y-axis shows the FL4-H fluorescence.

[0058] Fig. 5 is a FACS analysis of BHK cells transfected with CytTs-EGFP, CytTs-EGFP-OriP and CytTs-EGFP-OPE. BHK cells were transfected with

CytTs EGFP, CytTs EGFP-OriP or CytTs EGFP-OPE and selected with puromycin at a concentration 5ug/ml. Two weeks after transfection cell populations were assayed for EGFP expression. Non-induced cells (37°C) and cells induced for 3 days at 29°C were trypsinized and analyzed by flow cytometry. GFP expression is shown on the X-axis (FL1-H) the Y-axis shows the FL4-H fluorescence.

[0059] Figs. 6A-6B demonstrate IFN $\beta$  expression in BHK cells transfected with CytTs-IFN $\beta$ -OPE. BHK cells were transfected with CytTs-IFN $\beta$ -OPE and selected with puromycin as described in the above. BHK-CytTs-IFN $\beta$ -OPE cell populations were then either kept at 37°C or shifted to 29°C for induction. Samples were taken from uninduced cells and from the induced cells 3, 4, 7 and 9 days after induction and quantitated by ELISA. The interferon concentration as international units (IU)/ml at the different time points is shown in panel A. 0 corresponds to uninduced control cells. The fold induction observed is shown in panel B.

[0060] Fig. 7 shows SEAP expression in BHK cells transfected with CytTs-SEAP-OPE. BHK cells were transfected with CytTs-SEAP-OPE and selected with puromycin as described above. BHK-CytTs-SEAP-OPE cell populations were then either kept at 37°C or shifted to 29°C for induction. Samples were taken from the uninduced cells and from the induced cells 3, 4, 7 and 9 days after induction and quantitated.

[0061] Fig. 8 demonstrates the improved EPO expression from pCytTS-OPE-EPO compared to pCytTS-OriP-EPO in 293 EBNA1 cells. Selected cell populations which had been transfected either with pCytTS-OPE-EPO or with pCytTS-OriP-EPO were either kept at 37°C or shifted to 29°C in order to induce the Sindbis replicon. Samples were taken from the induced and uninduced cells 0, 4 and 8 days after induction and were quantified by ELISA. The EPO concentration at the different time points is shown in mg/l on the Y-axis.

[0062] Fig. 9 shows EPO expression in 293EBNA cells transfected with vector pCytTS-EPO-OriP. Selected cell populations were either kept at 37°C

or shifted to 29°C in order to induce the Sindbis replicon. Cell culture supernatant was collected from uninduced and induced cells seven days after induction. EPO levels were quantified by ELISA.

## DETAILED DESCRIPTION OF THE INVENTION

### OVERVIEW

- [0063] The present invention is directed to compositions and methods that are useful, *e.g.*, in the production of polypeptides and/or untranslated RNA molecules in host cells. The compositions and methods of the invention are also useful, *e.g.*, for the isolation and/or purification of polypeptides, polypeptides and untranslated RNA molecules that are produced in host cells.
- [0064] According to one aspect of the invention, nucleic acid molecules are provided. The nucleic acid molecules of the invention comprise: (A) a first polynucleotide element which encodes an RNA molecule, said RNA molecule comprising: (a) at least one *cis*-acting sequence element; (b) a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and (c) at least one second nucleotide sequence selected from the group consisting of: (i) a second open reading frame encoding a polypeptide; (ii) a nucleotide sequence complementary to all or a part of the second open reading frame of (i); and (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof; wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase; (B) a second polynucleotide element comprising an origin of replication; and (C) a third polynucleotide element encoding a replication initiation factor capable of recognizing said origin of replication.
- [0065] According to another aspect of the invention, expression systems are provided. The expression systems of the invention comprise one or more nucleic acid molecules, wherein said one or more nucleic acid molecules comprise: (A) a first polynucleotide element which encodes an RNA

molecule, said RNA molecule comprising: (a) at least one *cis*-acting sequence element; (b) a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and (c) at least one second nucleotide sequence selected from the group consisting of: (i) a second open reading frame encoding a polypeptide; (ii) a nucleotide sequence complementary to all or a part of the open reading frame of (i); and (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof; wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase; (B) a second polynucleotide element comprising an origin of replication; and (C) a third polynucleotide element encoding a replication initiation factor capable of recognizing said origin of replication.

**[0066]** In the expression systems of the invention, said first, second and third polynucleotide elements may each be on a separate nucleic acid molecule. Alternatively, said first, second and third polynucleotide elements may each be on a single nucleic acid molecule. In certain embodiments, said first and second polynucleotide elements will be on a single nucleic acid molecule while said third polynucleotide element is on a different nucleic acid molecule. In certain other embodiments, said first and third polynucleotide elements will be on a single nucleic acid molecule, while said second polynucleotide element is on a different nucleic acid molecule. In certain other embodiments, said second and third polynucleotide elements will be on a single nucleic acid molecule, while said first polynucleotide element is on a different nucleic acid molecule.

**[0067]** According to another aspect of the invention, recombinant host cells and *in vitro* cell cultures comprising recombinant host cells are provided. The recombinant host cells of the invention comprise: (A) a first polynucleotide element which encodes an RNA molecule, said RNA molecule comprising: (a) at least one *cis*-acting sequence element; (b) a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and (c) at



least one second nucleotide sequence selected from the group consisting of: (i) a second open reading frame encoding a polypeptide; (ii) a nucleotide sequence complementary to all or a part of the second open reading frame of (i); and (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof; wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase; (B) a second polynucleotide element comprising an origin of replication; and (C) a third polynucleotide element encoding a replication initiation factor capable of recognizing said origin of replication.

[0068] According to another aspect of the invention, methods are provided for producing, or regulating the expression of, a polypeptide or untranslated RNA molecule. The methods according to this aspect of the invention comprise: (a) introducing a nucleic acid molecule or an expression system of the invention into a host cell to produce a recombinant host cell; and (b) culturing said recombinant host cell. The methods of the invention may further comprise recovering said polypeptide or untranslated RNA molecule.

[0069] According to another aspect of the invention, kits are provided comprising the nucleic acid molecules, expression systems and/or host cells of the invention.

[0070] In certain embodiments, said second open reading frame that is included within the compositions and methods of the invention is in a translatable format after one or more RNA-dependent RNA replication events. The expression "a translatable format" is intended to mean an RNA from which a protein can be made.

[0071] The RNA-dependent RNA polymerase that is included in, or used with, the compositions and methods of the invention may, in certain embodiments, be selected from the group consisting of: (a) a temperature-sensitive RNA-dependent RNA polymerase, (b) a non-cytopathic RNA-dependent RNA polymerase, and (c) a temperature-sensitive, non-cytopathic RNA-dependent RNA polymerase.

[0072] When the RNA-dependent RNA polymerase that is included in, or used with, the compositions and methods of the invention is temperature

sensitive, the temperature-sensitive RNA-dependent RNA polymerase may have replicase activity at temperatures below 34°C which is at least five fold, at least ten fold, at least twenty fold, at least thirty fold, at least forty fold, at least fifty fold, at least one hundred fold, or at least one thousand fold greater than the replicase activity exhibited at 34°C or above. For example, the temperature-sensitive RNA-dependent RNA polymerase may have replicase activity at 34°C that is at least five fold greater than the replicase activity exhibited at 29°C. The temperature-sensitive RNA-dependent RNA polymerase may, in certain embodiments, exhibit no or undetectable replicase activity (using methods of detection known in the art) at 34°C or above.

[0073] The RNA-dependent RNA polymerase that is included in, or used with, the compositions and methods of the invention may, in certain embodiments, be of viral origin. For example, the RNA-dependent RNA polymerase may be of alphaviral origin. In specific embodiments of the invention, the RNA-dependent RNA polymerase is derived from a Sindbis virus, a Semliki Forest virus or an Aura virus. The RNA-dependent RNA polymerase may be derived from a virus selected from the group consisting of Bebaru virus, Cabassou virus, Chikungunya virus, Easter equine encephalomyelitis virus, Fort Morgan virus, Getah virus, Kyzylagach virus, Mayoaro virus, Middleburg virus, Mucambo virus, Ndumu virus, Pixuna virus, Tonate virus, Trinita virus, Una virus, Western equine encephalomyelitis virus, Whataroa virus, Venezuelan equine encephalomyelitis virus (VEE), and Ross River virus.

[0074] The origin of replication that is included in, or used with, the compositions and methods of the invention may be derived from a prokaryotic organism, a eukaryotic organism (e.g., a yeast, insect or mammal), or a virus. For example, the origin of replication may be derived from a DNA virus, e.g., a DNA virus that allows for episomal replication. The origin of replication may be derived from a DNA virus selected from the group consisting of Herpesvirus, Epstein-Barr virus (EBV), Papillomavirus, Polyomavirus, Adenovirus, and Hepadnavirus. In a specific embodiment, the origin of replication is *oriP*, derived from EBV.

- [0075] The replication initiation factor that is included in, or used with, the compositions and methods of the invention is, in certain embodiments, capable of operating as a plasmid maintenance factor. As used herein, the term "plasmid maintenance factor" is intended to mean a factor, which supports the distribution of the episomes to the daughter cells upon cell division. In the case of EBNA-1, this is achieved by the binding of the protein to the episome and, at the same time, to the chromosome, thereby leading to proper distribution of the episomes to the daughter cells. The replication initiation factor may be derived from a prokaryotic organism, a eukaryotic organism (e.g., a yeast, insect or mammal), or a virus. For example, the replication initiation factor may be derived from a DNA virus, e.g., a DNA virus that allows for episomal replication. The replication initiation factor may be derived from a DNA virus selected from the group consisting of Herpesvirus, Epstein-Barr virus (EBV), Papillomavirus, Polyomavirus, Adenovirus, and Hepadnavirus. In a specific embodiment, the replication initiation factor is the EBNA-1 protein, derived from EBV.
- [0076] The origin of replication and the replication initiation factor that are included in, or used with, the compositions and methods of the invention may, in certain embodiments, be derived from the same organism or the same virus. Alternatively, the origin of replication and the replication initiation factor may be derived from different organisms or viruses.
- [0077] The nucleic acid molecules and expression systems of the invention may further comprise a fourth polynucleotide element, wherein said fourth polynucleotide element comprises a selection marker. Any selection marker known in the art may be used, including, e.g., markers that confer resistance to antibiotics. An exemplary selection marker is one that confers resistance to puromycin. Alternatively, selection markers may be used that confer resistance to hygromycin, gpt, neomycin, zeocin, ouabain, blasticidin, or bleomycin, or markers such as DHFR, hisD, trpB, or glutamine synthetase.
- [0078] The second open reading frame that is included in, or used with, the compositions and methods of the invention may encode, in certain embodiments, any polypeptide of interest. For example, the second open

reading frame may encode, *e.g.*, a cytokine, a lymphokine, a tumor necrosis factor, an interferon (*e.g.*, beta-interferon, including human beta-interferon), a toxic polypeptide, or a prodrug.

[0079] The second open reading frame that is included in, or used with, the compositions and methods of the invention may encode, in certain embodiments, an untranslated RNA molecule. Exemplary untranslated RNA molecules include, *e.g.*, an antisense RNA molecule, a tRNA molecule, an rRNA molecule and a ribozyme. Other useful untranslated RNA molecules are known in the art.

[0080] Thus, in certain exemplary embodiments, the invention provides compositions and methods which allow for the production of high amounts of specific RNA molecules produced in host cells upon their transient transfection with said polynucleotides. This high production results, *e.g.*, from the use of a self-replicating episomal DNA plasmid combined with an RNA-dependent RNA polymerase, which amplifies the mRNA of interest.

[0081] Expression, using the compositions and methods of the invention, may be regulated precisely, if necessary. Such precise regulation can be achieved, e.g., from the use of a temperature-sensitive RNA-dependent RNA polymerase which only replicates RNA molecules to form additional RNA molecules at permissive temperatures.

[0082] According to the present invention, polypeptides and/or untranslated RNA molecules may be produced in transient transfection experiments due to the use of nucleic acid molecules comprising an origin of replication and DNA encoding a protein (a factor or a DNA-dependent replicase) that can recognize the origin of replication, thereby resulting in replication of the plasmid.

[0083] The expression systems of the invention, in certain embodiments, comprise alphavirus DNA vectors that can be used to create transiently transfected cell lines. Said alphavirus DNA vectors may carry genes encoding a non-cytopathic replicase and viral non-structural proteins, thereby being able to produce high amounts of polypeptide. An induction of expression is achieved when the activity of the temperature-sensitive replicase is switched on by reducing the incubation temperature of the transfected cells from 37°C

to a temperature lower than 34°C. Protein expression in the host at 37°C is, preferably, below the level of detection.

## DEFINITIONS

[0084] The following definitions are provided to clarify the subject matter, which the inventors consider to be the present invention.

[0085] As used herein, a nucleic acid molecule is a sequence of contiguous nucleotides (riboNTPs, dNTPs or ddNTPs, or combinations thereof) of any length which may encode a full-length polypeptide or a fragment of any length thereof, or which may be non-coding. As used herein, the terms "nucleic acid molecule," and "polynucleotide," "polynucleotide construct," and "polynucleotide element" may be used interchangeably.

[0086] As used herein, the term "alphavirus" refers to any of the RNA viruses included within the genus *Alphavirus*. Descriptions of the members of this genus are contained in Strauss and Strauss, *Microbiol. Rev.*, 58:491-562 (1994). Examples of alphaviruses are selected from the group comprising Aura virus, Bebaru virus, Cabassou virus, Chikungunya virus, Easter equine encephalomyelitis virus, Fort morgan virus, Getah virus, Kyzylagach virus, Mayoaro virus, Middleburg virus, Mucambo virus, Ndumu virus, Pixuna virus, Tonate virus, Trinita virus, Una virus, Western equine encephalomyelitis virus, Whataroa virus, Sindbis virus (SIN), Semliki forest virus (SFV), Venezuelan equine encephalomyelitis virus (VEE), and Ross River virus.

[0087] As used herein, the term "purified" used in reference to a molecule means that the concentration of the molecule being purified has been increased relative to molecules associated with it in its natural environment. Naturally associated molecules include polypeptides, nucleic acids, lipids and sugars but generally do not include water, buffers, and reagents added to maintain the integrity or to facilitate the purification of the molecule. For example, even if mRNA is diluted with an aqueous solvent during oligo dT column chromatography, mRNA molecules are purified by this chromatography if

naturally associated nucleic acids and other biological molecules do not bind to the column and are separated from the subject mRNA molecules.

[0088] As used herein, the term "isolated" used in reference to a molecule means that the molecule has been removed from its native environment. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated." Further, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Isolated RNA molecules include *in vivo* or *in vitro* RNA replication products of DNA and RNA molecules. Isolated nucleic acid molecules further include synthetically produced molecules. Additionally, vector molecules contained in recombinant host cells are also isolated. Thus, not all "isolated" molecules need to be "purified."

[0089] As used herein, the phrase "low or undetectable," when used in reference to gene expression level, refers to a level of expression which is either significantly lower than that seen when the gene is maximally induced (e.g., at least five fold lower) or is not readily detectable by the methods used in the following examples section.

[0090] As used herein, the phrase "individual" refers to multicellular organisms and includes both plants and animals. Preferred multicellular organisms are animals, more preferred are vertebrates, even more preferred are mammals, and most preferred are humans.

[0091] As used herein, the term "pCytTs" means any temperature sensitive, non-cytopathic inducible alphaviral expression system containing a functional promoter to drive the transcription of mRNA from the nucleic acid molecule of the invention. Examples for such expression systems or vectors may be, e.g., the vector pCytTs (SEQ ID NO: 1) containing a RSV promoter or the vector pCytTs2.1 (SEQ ID NO: 2) containing a CMV promoter. When combinatorial constructs with pCytTs are mentioned in the Examples, they generally contain a CMV promoter, even if the abbreviation of the combinatorial constructs does - for the sake of simplicity - only refer to pCytTs and not to pCytTs2.1.

- [0092] As used herein, the phrase "*cis*-acting" sequence refers to nucleic acid sequences to which a replicase binds to catalyze the RNA-dependent replication of RNA molecules. These replication events result in the replication of the full-length and partial RNA molecules and, thus, the alphavirus subgenomic promoter is also a "*cis*-acting" sequence. *Cis*-acting sequences may be located at or near the 5' end, 3' end, or both ends of a nucleic acid molecule, as well as internally.
- [0093] As used herein, the phrase "RNA-dependent RNA polymerase" refers to a polymerase which catalyzes the production of an RNA molecule from another RNA molecule. This term is used herein synonymously with the term "replicase."
- [0094] As used herein, the phrase "non-infective packaged RNA molecules" refers to packaged RNA molecules which will essentially undergo only one round of host cell infection and are not pathogenic. These molecules are thus "infective" but only for a single infectious entry into a host cell and are not capable of reproducing to form additional infectious particles.
- [0095] As used herein, the term "transcription" refers to the production of RNA molecules from DNA templates catalyzed by RNA polymerases.
- [0096] As used herein, the phrase "RNA-dependent RNA replication event" refers to processes which result in the formation of an RNA molecule using an RNA molecule as a template.
- [0097] As used herein, the term "vector" refers to an a polynucleotide construct, typically a plasmid or a virus, used to transmit genetic material to a host cell. Preferably, the term "vector" as used herein refers to an agent such as a plasmid, and even more preferably to a circular plasmid, A vector as used herein may be composed of either DNA or RNA. Preferably, a vector as used herein is composed of DNA.
- [0098] As used herein, the term "heterologous sequence" refers to a second nucleotide sequence present in a vector of the invention. The term "heterologous sequence" also refers to any amino acid or RNA sequence encoded by a heterologous DNA sequence contained in a vector of the invention. Heterologous nucleotide sequences can encode polypeptides or

RNA molecules normally expressed in the cell type in which they are present or molecules not normally expressed therein (*e.g.*, Sindbis structural proteins).

[0099] As used herein, the phrase "untranslated RNA" refers to an RNA sequence or molecule which does not contain an open reading frame or encodes an open reading frame, or portion thereof, but in a format in which an amino acid sequence will not be produced (*e.g.*, no initiation codon is present). Examples of such molecules are tRNA molecules, rRNA molecules, and ribozymes. Antisense RNA may be untranslated but, in some instances antisense sequences can be converted to a translatable sense strand from which a polypeptide is produced.

[00100] As used herein, the phrase "temperature-sensitive" refers to an enzyme which readily catalyzes a reaction at one temperature but catalyzes the same reaction slowly or not at all at another temperature. An example of a temperature-sensitive enzyme is the replicase protein encoded by the pCytTs vector, which has readily detectable replicase activity at temperatures below 34°C and has low or undetectable activity at 37°C.

[00101] As used herein, the phrase "permissive temperature" refers to temperatures at which an enzyme has relatively high levels of catalytic activity.

[00102] As used herein, the phrase "restrictive temperature" refers to temperatures at which an enzyme has low or undetectable levels of catalytic activity. Both "hot" and "cold" sensitive mutants are known and, thus, a restrictive temperature may be higher or lower than a permissive temperature.

[00103] As used herein, the term "recombinant host cell" refers to a host cell into which one or more nucleic acid molecules of the invention have been introduced.

[00104] As used herein, the term "replication initiation factor" refers to a protein or a DNA sequence encoding such a protein that is able to recognize an origin of replication leading to replication of plasmids containing this origin of replication.

[00105] As used herein, the term "origin of replication" refers to a DNA sequence that is recognized by a replication initiation factor or a DNA



-27-

replicase leading to replication of a plasmid containing the origin of replication. The expression "recognized by a replication initiation factor" is intended to mean that a replication initiation factor can physically interact with all or a portion of an origin of replication sequence, thereby causing or stimulating molecular mechanisms that ultimately cause all or a portion of the DNA molecule comprising the origin of replication to be replicated.

[00106] As used herein the term "inducible" means that in an inactive state, the basal activity of a system should be minimal, while in the active state high levels of gene expression should be rapidly inducible.

[00107] When the terms "one," "a," or "an" are used in this disclosure, they mean "at least one" or "one or more," unless otherwise indicated.

## NUCLEIC ACID MOLECULES AND EXPRESSION SYSTEMS

[00108] Certain preferred embodiments of the present invention are based on the vectors used in the previously described pCytTs expression system (WO 99/50432, the disclosure of which is hereby incorporated by reference in its entirety). These vectors are constitutively transcribed in host cells to produce mRNA molecules having two open reading frames. These open reading frames, which may or may not be produced from the same nucleic acid molecule, encode a temperature-sensitive replicase and a heterologous gene of interest. The first open reading frame is translated to produce an RNA-dependent RNA polymerase. The second open reading frame, encoding all or part of one or more polypeptides of interest, is not translated until after at least one RNA-dependent RNA replication event.

[00109] The nucleic acid molecule of the present invention comprise, *inter alia*, at least one second nucleotide sequence. The second nucleotide sequence is preferably operably linked a promoter which is recognized by an RNA-dependent RNA polymerase. The second nucleotide sequence, in certain embodiments, may comprise a second open reading frame, or other nucleotide sequence of interest, as well as other elements. For instance, the second nucleotide sequence may comprise a 5' promoter which is capable of initiating

synthesis of RNA *in vivo*, 5' and/or 3' sequences enabling replication of the RNA molecule (5' and 3' *cis* acting sequence elements), as well as a sequence of interest which is translatable only after at least one replication event. Replication is catalyzed by an RNA-dependent RNA polymerase which is encoded alternatively on the same or on a different mRNA molecule. The sequence of interest may be encoded in sense, plus (+), orientation downstream of a viral RNA promoter. Translation of the coding sequence of the gene of interest is inhibited by a 5' sequence which, in the case of the single-vector system, will generally be the replicase sequence. In the multiple-vector system, a 5' sequence can inhibit translation by having one or more short open reading frames with associated stop codons which lead to the detachment of ribosomes. Similarly, any sequence which inhibits the traveling or binding of ribosomes to the sequence of interest can be used as a 5' sequence which inhibits translation (Voet and Voet, BIOCHEMISTRY, John Wiley & Sons, Inc. (1990)).

[00110] Another method for preventing translation of nucleotide sequences in most biological systems involves the insertion of the sequence in an antisense direction. This method of inhibiting translation is based on the principle that translation will generally only occur after the replication of this minus (-) strand RNA into a plus strand having an open reading frame in a sense orientation. The translated sense strand is formed by RNA replication and serves as a template for ribosomes and protein synthesis. As shown previously (WO 99/50432), production of amino acid sequences can occur even when the gene of interest is inserted into the DNA molecule in an orientation which will result in the formation of antisense RNA sequence 3' to the subgenomic promoter. Thus, the second open reading frame may also comprise a sequence complementary to all or part of the second open reading frame described above and expression of the encoded amino acid sequence will still occur. When the production of an untranslated antisense RNA sequence is desired, the RNA molecule can be designed so that it will not serve as a template for protein synthesis. For example, the RNA can be designed so that an initiation codon is not present.

**[00113]** The nucleic acid molecules and expression systems of the invention also comprise a third polynucleotide element encoding a replication initiation factor. The replication initiation factor will be capable of recognizing the

origin of replication that is included within the nucleic acid molecule or expression system. The third polynucleotide element of the invention may encode, *e.g.*, a replication initiation factor derived from prokaryotic organism, eukaryotic organism (*e.g.*, yeasts, insects, and/or mammals) and/or viruses. For example, a replication initiation factor that may be used with the present invention is one that is derived from a DNA virus such as, *e.g.*, a Herpesvirus, a Papillomavirus, an Adenovirus, or a Hepadnavirus. In certain embodiments, the replication initiation factor that is used with the invention is capable of operating as a plasmid maintenance factor. In a preferred embodiment, the replication initiation factor is derived from Epstein-Barr virus (EBV), such as, *e.g.*, the EBNA-1 protein, or a portion thereof.

[00114] The replication initiation factor (*e.g.*, a nucleotide sequence that encodes a replication initiation factor) and the origin of replication that are included within the nucleic acid molecules and expression systems of the present invention may be derived from the same organism or from the same virus. Alternatively, the replication initiation factor and the origin of replication may be derived from different organisms or from different viruses.

[00115] In an exemplary embodiment of the invention, nucleic acid molecules are provided that are constructed by introducing defined Epstein-Barr virus (EBV) sequences into vectors such as, *e.g.*, those included within the pCytTs system. The resulting exemplary nucleic acid molecules of the invention are capable of replicating as non-integrated autonomous episomal molecules in the transformed host cells.

[00116] The defined sequences, which may be introduced into the vectors of the pCytTs system, comprise, *e.g.*, either one or two elements of Epstein-Barr virus, (OriP alone or together with EBNA-1 gene) which permit plasmid maintenance. The origin of replication OriP is a cis-acting sequence and needs to be inserted into the plasmid vector sequence. Plasmids containing this origin of replication are able to be maintained in cells expressing the replication initiation factor which recognizes the origin of replication; one of these factors is EBNA-1. EBNA-1 gene function can be provided in cis by introducing the sequence into the same plasmid or can be provided in trans by

co-transfection with a second replicating or non-replicating plasmid as well as by providing it from stably transduced cell lines expressing EBNA-1 gene from an integrated copy. Furthermore, in certain embodiments, more than one copy of the sequence expressing the replication initiation factor, *e.g.* EBNA-1, is provided. For example, EBNA-1 gene function may be provided on one or more plasmids, wherein none, some or all of those plasmids may additionally comprise the origin of replication, and which plasmids are used for transfecting cell lines, preferably cell lines that have the sequence of EBNA-1 stably integrated within its genome. The presence of both OriP and EBNA1 sequences in the same plasmid renders replication less dependent on the host cell type. (Reisman, D. *et al.*, *Mol. Cell Biol.* 5: 1822-1832 (1985); Yates, J.L. *et al.*, *Nature* 313:812-815 (1985); U.S. Patent No. 4,686,186). U.S. Patent No. 4,686,186 describes the transfection of cells with a single plasmid containing the EBV OriP, the EBNA-1 gene and a gene encoding a protein of interest.

[00117] Although such cell lines have the advantage of stable long-term expression of the replication initiation factor and durable support of replication and maintenance of OriP containing plasmids, there are not many EBNA-1 expressing cell lines commercially available (ATCC : 293HEK-EBNA1 and CV1-EBNA1). Alternatively, plasmids that already carry the EBNA-1 gene and the gene of interest in *cis* on the same episome are used to transfect cells and commercial vectors such as pCEP4 (Invitrogen) are available. However, current vectors designed for constitutive expression of polypeptides carrying OriP and EBNA-1 on the episomal construct in *cis* and wherein the expression is neither regulatable or inducible, are not applicable in cases where cell-toxic polypeptides need to be expressed.

[00118] To our knowledge, an inducible expression system having the features of a self-replicating DNA episome and that allows the expression of a polypeptide or untranslated RNA sequence of interest from a self-replicating RNA molecule, triggered by a simple temperature shift, has not been described.

[00119] Certain embodiments of the invention are directed to nucleic acid molecules which are transcribed to produce a mRNA molecule having two open reading frames, *e.g.*, one open reading frame which encodes a replicase, and another open reading frame which encodes a nucleotide sequence of interest. The nucleic acid molecules of the invention may contain a promoter sequence which drives transcription to produce mRNA molecules having coding sequences of both open reading frames. The mRNA sequences of the first open reading frame are preferably translated to produce a replicase required for the expression of the RNA sequences of the second open reading frame. The second open reading frame preferably encodes one or more polypeptides of interest. In addition, the vectors contain an origin of replication and a DNA polymerase recognizing the origin of replication.

**[00120]** Further, once the first mRNA molecule has been transcribed from the DNA vector, additional RNA-dependent RNA replication events may occur to amplify the first mRNA sequence and to produce RNA molecules with strand polarity which is the opposite of the first mRNA sequence.

**[00121]** The second open reading frame of the nucleic acid molecules of the invention will preferably only be expressed after partial replication of a full-length RNA molecule. This partial replication of the full-length RNA molecules is driven by a promoter sequence composed of RNA (*e.g.*, an alphaviral subgenomic promoter sequence). In addition, the plasmid encoding the replicase and the gene of interest will preferably replicate during division of the host cell due to the presence of origin of replication and a protein factor recognizing the origin of replication.

**[00122]** While the gene of interest may be encoded by the same RNA molecule as the RNA-dependent RNA replicase protein, this gene may also be encoded by a separate RNA molecule. Thus, the invention further provides both single- and multiple-vectors systems for expressing a gene of interest. Moreover, while all plasmids used preferably contain an origin of replication, the gene encoding the replication initiation factor may be on a separate plasmid or stably integrated into the genome of the host cell.

**[00123]** In a single-vector system of the invention, sequences encoding the first open reading frame, the second nucleotide sequence are components of the same nucleic acid molecule, the origin of replication and the initiation factor. Thus, all of the components required for regulated expression of the gene of interest and plasmid amplification are contained in a single nucleic acid molecule (*i.e.*, DNA or RNA).

**[00124]** In a multiple-vector system of the invention, sequences encoding the first open reading frame, or sub-portions thereof, and the second nucleotide sequence are components of different nucleic acid molecules. These multiple-vector systems thus may comprise two or more nucleic acid molecules. For example, the replicase and the gene of interest can each be encoded by different nucleic acid molecules. However, all vectors may contain an origin of replication.

**[00125]** While any functional promoter can be used to drive the transcription of mRNA from the nucleic acid molecule of the invention, the promoter is preferably a constitutive RNA polymerase II promoter (*e.g.*, Rous Sarcoma Virus (RSV), cytomegalovirus (CMV), simian virus 40 (SV40), myeloproliferative sarcoma virus (MPSV), glucocorticoid, metallothionein, *Herpes simplex* virus thymidine kinase (HSVTK), human immune deficiency (HIV), mouse mammary tumor virus (MMTV), human polyomavirus BK (BKV), or Moloney murine leukemia virus (MuLV) promoter). Additional promoters suitable for use in the practice of the present invention are known in the art (*see, e.g.*, Lee, A. *et al.*, *Mol. Cells*. 7:495-501 (1997)).

**[00126]** The nucleic acid molecules of the invention may further comprise a fourth polynucleotide element, wherein said fourth polynucleotide element comprises a selection marker. The selection marker may facilitate the cloning and amplification of the vector sequences in prokaryotic and eukaryotic organisms. In certain embodiments, the selection marker will confer resistance to a compound or class of compounds, such as an antibiotic. An exemplary selection marker that can be used with the nucleic acid molecules and expression systems of the present invention is one that confers resistance to puromycin. Alternatively, selection markers may be used that confer

resistance to hygromycin, gpt, neomycin, zeocin, ouabain, blasticidin, or bleomycin, or markers such as DHFR, hisD, trpB, or glutamine synthetase.

[00127] The pCytTs vector contains an ampicillin resistance marker for positive selection in bacterial host cells and an *E. coli* origin of replication (*i.e.*, ColE1). A considerable number of sequences encoding additional selection markers and origins of replication are known in the art (*see, e.g.*, Sambrook, J. *et al.*, eds., MOLECULAR CLONING, A LABORATORY MANUAL, 2nd. edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel, F. *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John H. Wiley & Sons, Inc. (1997)).

[00128] The nucleic acid molecules and expression systems of the invention comprise a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding a replicase, *e.g.*, an RNA-dependent RNA polymerase. The first nucleotide sequence may, in certain embodiments, also comprise 5' and 3' *cis*-acting sequences, and junction sequences containing a subgenomic promoter.

[00129] The RNA-dependent RNA polymerases that can be used with the present invention may be of viral origin. The RNA-dependent RNA polymerase may be derived, *e.g.*, from virus selected from the group consisting of: Bebaru virus, Cabassou virus, Chikungunya virus, Easter equine encephalomyelitis virus, Fort morgan virus, Getah virus, Kyzylagach virus, Mayoaro virus, Middleburg virus, Mucambo virus, Ndumu virus, Pixuna virus, Tonate virus, Trinita virus, Una virus, Western equine encephalomyelitis virus, Whataroa virus, Venezuelan equine encephalomyelitis virus (VEE), and Ross River virus. In a preferred embodiment, the RNA-dependent RNA polymerase of the invention is derived from an alphavirus such as, *e.g.*, a Sindbis virus, a Semliki Forest virus, and/or an Aura virus.

[00130] In certain embodiments of the invention, the RNA-dependent RNA polymerase that is encoded by the first open reading frame of the nucleic acid molecules of the invention will be a non-cytopathic RNA-dependent RNA polymerase. The RNA-dependent RNA polymerase may, alternatively, be a temperature sensitive RNA-dependent RNA polymerase. In other



embodiments, the RNA-dependent RNA polymerase may be a non-cytopathic, temperature-sensitive RNA-dependent RNA polymerase.

[00131] When the RNA-dependent RNA polymerase that is encoded by the first open reading frame of the nucleic acid molecules of the invention is a temperature-sensitive RNA-dependent RNA polymerase, the RNA-dependent RNA polymerase may, *e.g.*, have replicase activity at temperatures below 34°C and low or undetectable replicase activity at 34°C or above. In certain embodiments, the temperature sensitive RNA-dependent RNA polymerase will be one that has replicase activity at 34°C which is at least five fold lower than the replicase activity exhibited at 29°C.

[00132] Preferred mutations which confer a non-cytopathic phenotype are those that are in the nsp2 gene (*e.g.*, the proline residue at position 726 is replaced with a serine or a leucine residue). Mutations are known in the art which render the replicase protein non-cytopathic (Weiss *et al.*, *J. Virol.* 33:463-474 (1980); Dryga *et al.*, *Virology* 228:74-83 (1997); Agapov *et al.*, *PNAS* 95:12989-12994 (1998)). In addition, mutations may be introduced into the nuclear localization signal of an alphaviral replicase which diminish toxicity of the replicon by, *e.g.*, modifying the distribution of the replicase-encoding sequence between the cytoplasm and the nucleus. (Rikkonen *et al.*, *Virology* 218:352-361 (1996)).

[00133] Mutations which render a replicase non-cytopathic may be introduced by a number of means, including site directed mutagenesis. Additional mutations that render an RNA-dependent RNA polymerase non-cytopathic can be identified by, *e.g.*, random mutagenesis and screening using methods and techniques that are well-known in the art. As will be understood by persons having ordinary skill in the art, the identification of additional mutations that render an RNA-dependent RNA polymerase non-cytopathic, temperature sensitive, or both, does not require an ability to predict the structural or functional consequences of any particular mutation.

[00134] About 20 years ago Weiss *et al.* (Weiss, B. *et al.*, *J. Virol.* 33:463-474 (1980)) established a persistently infected culture of BHK cells. The mutation responsible for this phenotype has been recently identified (Dryga, S.A. *et al.*,

*Virology* 228:74-83 (1997)). Another mutation allowing the regulation of the mRNA transcription via temperature shifts was identified by Burge and Pfefferkorn (Burge, B.W. & Pfefferkorn, E.R., *Virology* 30:203-214 (1966)) and described in more detail by Xiong *et al.* (Xiong, C. *et al.*, *Science* 243:1188-1191 (1989)).

[00135] Temperature sensitivity (*ts*) may be conferred, for example, by the introduction of a mutation in the *nsp4* gene of the replicase. Preferably, mutations which confer a temperature-sensitive phenotype upon replicase activities are in a protein in complementation group F (Lemm *et al.*, *J. Virol.* 64:3001-3011 (1990)). For example, a temperature-sensitive phenotype may be conferred by changing Gly 153 of *nsp4* to Glu. Additionally, any other mutation which renders replicase activity temperature-sensitive can be used in the practice of the invention. Methods for creating and identifying new temperature-sensitive mutants are described by Pfefferkorn (Burge and Pfefferkorn, *Virol.* 30:204-213(1966); Burge and Pfefferkorn, *Virol.* 30:214-223 (1966)). Other methods will be appreciated by those of ordinary skill in the art.

[00136] While most temperature-sensitive mutants are "hot" sensitive, "cold" sensitive mutations are also known (*see, e.g.*, Schwer, B. *et al.*, *Nucleic Acids Res.* 26:803-809 (1998), Mathe, E. *et al.*, *J. Cell Sci.* 111:887-896 (1998), Doedens, J. *et al.*, *J. Virol.* 71:9054-9064 (1997), Patterson, B. *et al.*, *J. Biol. Chem.* 272:27612-27617 (1997)). The temperature-sensitive replicase may be "cold" or "hot" sensitive and thus will catalyze RNA replication only at temperatures either above or below restrictive temperatures. In one embodiment, RNA replication occurs at detectable levels only at temperatures lower than 34°C. In contrast to all previously known regulatable DNA expression systems, the basal level of expression in recombinant host cells containing the pCytTs vector in the inactive state at 37°C is below the level of detection using standard methods (*e.g.*, those used in the following examples) even under condition of transient transfection. This low level of expression is apparent from the data presented in Fig.4.

**[00137]** The production of additional temperature sensitive and/or non-cytopathic replicases has been described in the art. (Lundstrom *et al.*, *Gene Ther. Mol. Biol.* 4:23-31 (1999), Lundstrom *et al.*, *Histochem. Cell. Biol.* 115:83-91 (2001)). Moreover, with respect to alphaviruses, the results obtained in one type of alphavirus can often be used to predict the results obtained in other alphaviruses. That is, the identification of mutations that render a replicase temperature-sensitive and/or non-cytopathic in one alphavirus type may be used to identify or create a similar temperature-sensitive and/or non-cytopathic replicase in a different alphavirus. This ability is related, in part, to the high level of sequence identity observed among alphavirus genomes, especially in their replicase-encoding sequences. (Weaver *et al.*, *J. Virol.* 71:613-623 (1997), Smyth *et al.*, *J. Virol.* 71:818-823 (1997), Kuhn *et al.*, *J. Virol.* 70:7900-7909 (1996)).

**[00138]** In certain embodiments, the first nucleotide sequence (comprising a first open reading frame encoding an RNA-dependent RNA polymerase) and the second nucleotide sequence (comprising, *e.g.*, a second open reading frame encoding a polypeptide) are contained on two separate nucleic acid molecules. In such an instance, the second nucleotide sequence may carry both *cis*-acting sequences and a 5' region which inhibits translation of the sequence of interest. The first nucleotide sequence can also be encoded by a nucleic acid molecule which is different than the second nucleotide sequence. Replication and translation of the second nucleotide sequence in this multi-vectors system can be regulatable by temperature if a temperature sensitive RNA-dependent RNA polymerase is used.

**[00139]** Additional mutations that render an RNA-dependent RNA polymerase non-cytopathic, temperature-sensitive, or both, can be identified by, *e.g.*, random mutagenesis and screening using methods and techniques that are well-known in the art. As will be understood by persons having ordinary skill in the art, the identification of additional mutations that render an RNA-dependent RNA polymerase non-cytopathic, temperature sensitive, or both, does not require an ability to predict the structural or functional consequences of any particular mutation. The identification of an RNA-dependent RNA

polymerase that is non-cytopathic, temperature-sensitive, or both, only involves, *e.g.*, the generation of random mutations and the functional testing of the replicases that are encoded from the mutated nucleic acid molecules. Such screening methods are routine in the art.

[00140] The nucleic acid molecules and expression systems of the invention can be also used to express of more than one gene of interest. For example, recombinant host cells can be transfected with more than one nucleic acid molecule of the invention wherein one nucleic acid molecule encodes both the replicase and a polypeptide of interest and additional nucleic acid molecules may encode additional polypeptides of interest. Similarly, when mutations conferring non-cytopathicity and temperature sensitivity are both used, genes encoding polypeptides having suitable mutations (*e.g.*, Pro 726 Ser in nsp2 and Gly 153 Glu in nsp4) may be on separate nucleic acid molecules. Additional variations would be apparent to those skilled in the art.

[00141] The nucleic acid molecules of the invention may also contain packaging signals which direct the packaging of RNA molecules into viral particles. These RNA molecules can be packaged in the presence of wild-type virus or defective helper virus RNA. A significant improvement was made with the development of defective helper RNA molecules (Bredenbeek, P. *et al.*, *J. Virol.* 67:6439-6446 (1993)). These RNA molecules contain *cis*-acting sequences, required for replication of the full-length transcription product, and subgenomic RNA promoter sequences which drive the expression of the structural protein genes. For example, in cells containing both RNA molecules with packaging signals and the defective helper virus RNA, alphaviral non-structural proteins allow for replication and amplification of the defective helper virus RNA sequences which are translated to produce virion structural proteins. Since the helper virus RNA lacks packaging signals, these molecules are not packaged into assembled virions. Thus, virion particles produced in this way contain essentially only RNA sequences encoding the gene of interest and, generally, other sequences required for temperature-sensitive regulation of gene expression. These non-infective packaged RNA molecules do not contain sequences encoding virion structural

proteins and, thus, undergo only one round of host cell infection and are not pathogenic.

[00142] Non-infective packaged RNA molecules can be used to infect a culture of suitable host cells simply by adding the particles to culture medium containing these cells. The preparation of non-infective alphaviral particles is described in a number of sources, including "Sindbis Expression System", Version C, (Invitrogen Catalog No. K750-1).

[00143] One exemplary application of this system is directed to the temperature-dependent production of non-infective, packaged RNA molecules. These packaged RNA molecules may be produced by a number of means including using recombinant host cells containing two different nucleic acid molecules (*e.g.*, a nucleic acid molecule of the invention and a nucleic acid molecule encoding a helper virus RNA sequence). For example, one of these nucleic acid molecules will encode, *e.g.*, an RNA molecule which contains packaging signal sequences, sequences encoding a non-cytopathic, temperature-sensitive replicase, and the gene of interest. The other nucleic acid molecule will contain sequences encoding alphaviral structural proteins downstream from an alphavirus subgenomic promoter. Using such a system, viral particles containing only RNA molecules with packaging signals will be produced at permissive temperatures in recombinant host cells. This is so because alphaviral structural proteins will only be produced at a permissive temperature. Additional variations of the above would be apparent to one skilled in the art.

[00144] The second nucleotide sequence that is included within the nucleic acid molecules and expression systems of the invention may be, *e.g.*, a second open reading frame encoding a polypeptide. Such a second open reading frame may, alternatively or interchangeably, be referred to as a nucleotide sequence of interest. A wide variety of nucleotide sequences of interest can be expressed by the nucleic acid molecules and expression systems of the invention. These sequences include, but are not limited to, sequences encoding lymphokines, cytokines, toxins, enzymes, prodrug converting enzymes, antigens which stimulate immune responses, single chain antibodies, polypeptides which

stimulate or inhibit immune responses, tumor necrosis factors, and various proteins with therapeutic applications (e.g., growth hormones and regulatory factors).

[00145] In certain embodiments, the second open reading frame (the second nucleotide sequence) of the invention may encode, *e.g.*, a cytokine or lymphokine (*e.g.*,  $\beta$ -interferon). Hematopoiesis is regulated by lymphokines and cytokines which stimulate the proliferation and/or differentiation of various hemopoietic cells. Representative examples of cytokines and lymphokines include interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-3 (IL-3), interleukin-4 (IL-4), interleukin-5 (IL-5), interleukin-6 (IL-6), interleukin-7 (IL-7), interleukin-8 (IL-8), interleukin-9 (IL-9), interleukin-10 (IL-10), interleukin-11 (IL-11), interleukin-12 (IL-12), interleukin-13 (IL-13), interleukin-14 (IL-14), interleukin-15 (IL-15), interleukin-16 (IL-16), interleukin-17 (IL-17), granulocyte colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), macrophage colony stimulating factor (M-CSF), and interferons.

**[00146]** In certain embodiments, the second open reading frame (the second nucleotide sequence) of the invention may encode secreted enzymes (*e.g.*, secreted alkaline phosphatase), cytoplasmic enzymes (*e.g.*, green fluorescent protein), or any number of other proteins with therapeutic applications (*e.g.*, human insulin, human coagulation Factor VIII).

[00147] The nucleic acid molecules and expression systems of the invention can also be used to express heterologous sequence encoding cytotoxic polypeptides. Cytotoxic polypeptides act to directly or indirectly inhibit cellular growth or metabolism. Representative examples of toxins include *Shigella* toxin, ricin, *Diphtheria* toxin, *Cholera* toxin, *Pseudomonas* exotoxin A, and *Herpes simplex* virus thymidine kinase (HSVTK). Within other embodiments of this invention, the heterologous sequence encodes a prodrug converting enzyme. A prodrug converting enzyme activates a compound with little or no cytotoxicity into a toxic product. Representative examples are HSVTK, alkaline phosphatase, guanine phosphoribosyl

transferase, and penicillin-V amidase. Examples of both cytotoxic polypeptides and prodrug converting enzymes are discussed in numerous sources including PCT/US97/06010, EP 0716148, and WO 96/17072. In addition, a vast array of signaling molecules and membrane proteins are toxic if expressed at high levels. All these molecules may be suitable for expression using the system of the present invention.

[00148] The second nucleotide sequence of the nucleic acid molecules and expression systems of the invention may also be a nucleotide sequence encoding an untranslated RNA molecule or complement thereof. Exemplary untranslated RNA molecules include, *e.g.*, antisense sequences, RNase P targeted sequences which induce gene down-regulation, and ribozymes. Smith S. *et al.* (*J. Virol.* 71:9713-9721 (1997)) describes alphaviral vectors used to express ribozyme sequences.

[00149] The nucleic acid molecules and expression systems of the invention can also be used to express virtually any polypeptide, including ones which have not as yet been identified but are encoded by nucleotide sequences contained in, for example, cDNA libraries or host cell chromosomes. Example of such polypeptides include secreted proteins and proteins from various cellular compartments. Heterologous sequences expressed by the vectors of the invention can encode polypeptides and RNA molecules from non-human species (*e.g.*, other mammals, plants, fungi, bacteria or viruses). These heterologous sequences may further encode viral membrane proteins (*e.g.*, HIV gp160) or viral polyproteins (*e.g.*, Sindbis structural proteins).

[00150] Nucleotide sequences may be added to the nucleic acid molecules and vectors of the invention which result in the production of a fusion protein. For example, such sequences can encode amino acids sequences which are fused to a protein encoded by a gene of interest and confer one or more functional characteristics upon the expression product. These amino acid sequences include sequences which will target the gene product for export from the cell (*e.g.*, a secretory sequence) or to a subcellular compartment (*e.g.*, the nucleus). Such amino acid sequences further include sequences which facilitate purification (*e.g.*, a six His "tag"). Depending on the amino acid sequence and

the function imparted by the fused sequence, the added amino acid sequences may or may not be cleaved from the translation product.

[00151] Fusion proteins also include proteins which have domains or regions derived from various different proteins. Examples of such a fusion protein are those containing domain II of *Pseudomonas* exotoxin, a domain or amino acid sequence which has binding affinity for a cell surface receptor associated with a particular cell type, and another amino acid sequence with a preselected biological activity. Domain II of *Pseudomonas* exotoxin will translocate across cell membranes. Using this system, fusion proteins can be designed which will bind to specific cells types, will translocate across the cytoplasmic membranes of these cells, and will catalyze predetermined intracellular biological reactions. A system of this type is described in Pastan *et al.*, U.S. Patent No. 5,705,163. Methods for identifying amino acid sequences, which bind to specific cell types are described in Wu, A., *Nature Biotech.* 14:429-431 (1996).

[00152] The nucleic acid molecules and expression systems of the invention may, in certain embodiments, further comprise genetic elements which confer additional functional characteristics such as selection markers.

[00153] Markers for the selection of prokaryotic and eukaryotic cells containing vectors the present invention are well known in the art and include puromycin, tetracycline, ampicillin, neomycin, hygromycin, gpt, zeocin, ouabain, blasticidin, bleomycin, and kanamycin resistance. Alternatively, selection markers such as DHFR, hisD, trpB, or glutamine synthetase may be used. Nucleotide sequences which result in high copy number amplification are also known in the art and include the ColE1 sequence contained in the pCytTs vector.



**[00154]** The invention includes methods of making recombinant host cells and recombinant host cells produced using the methods of the invention. For example, the methods of the invention comprise introducing one or more nucleic acid molecules or expression systems described herein into a host cell. A variety of different recombinant host cells can be produced which contain the nucleic acid molecules and expression systems of the invention. Alphaviruses, for example, are known to have a wide host range. Sindbis virus, for example, infects cultured mammalian, reptilian, and amphibian cells, as well as some insect cells (Clark, H., *J. Natl. Cancer Inst.* 51:645 (1973); Leake, C., *J. Gen. Virol.* 35:335 (1977); Stollar, V. in *THE TOGAVIRUSES*, R.W. Schlesinger, Ed., Academic Press, (1980), pp.583-621).

**[00155]** Thus, numerous host cells can be used in the practice of the invention. Representative host cells that may be used with the invention include, but are not limited to, bacterial cells, yeast cells, plant cells and animal cells. Preferred bacterial host cells include *Escherichia* spp. cells (particularly *E. coli* cells and most particularly *E. coli* strains DH10B, Stbl2, DH5, DB3, DB3.1, DB4 and DB5), *Bacillus* spp. cells (particularly *B. subtilis* and *B. megaterium* cells), *Streptomyces* spp. cells, *Erwinia* spp. cells, *Klebsiella* spp. cells, *Serratia* spp. cells (particularly *S. marcescens* cells), *Pseudomonas* spp. cells (particularly *P. aeruginosa* cells), and *Salmonella* spp. cells (particularly *S. typhimurium* and *S. typhi* cells). Preferred animal host cells include insect cells (most particularly *Drosophila melanogaster* cells, *Spodoptera frugiperda* Sf9 and Sf21 cells and *Trichoplusia* High-Five cells), nematode cells (particularly *C. elegans* cells), avian cells, amphibian cells (particularly *Xenopus laevis* cells), reptilian cells, and mammalian cells (most particularly NIH3T3, CHO, COS, VERO, BHK, HEK, other rodent cells, and human cells). Preferred yeast host cells include *Saccharomyces cerevisiae* cells and *Pichia pastoris* cells. BHK, COS, Vero, HeLa and CHO cells are particularly suitable for the production of heterologous polypeptides because they have the potential to glycosylate heterologous proteins in a manner similar to human cells (Watson, E. *et al.*, *Glycobiology* 4:227, (1994)) and can be selected (Zang, M. *et al.*, *Bio/Technology* 13:389 (1995)) or genetically engineered

(Renner W. *et al.*, *Biotech. Bioeng.* 47:476 (1995); Lee K. *et al.* *Biotech. Bioeng.* 50:336 (1996)) to grow in serum-free medium, as well as in suspension.

**[00156]** Example 2 shows that the expression system of the present invention can function in BHK cells and 293HEK cells. Moreover, we have previously demonstrated that the pCytTs system works efficiently in additional cells lines such as CHO-K1 and COS-7 (Boorsma *et al.*, *Nature Biotech.* 18:429-432 (2000)).

**[00157]** The nucleic acid molecules and/or expression systems of the invention may be introduced into host cells using well known techniques of infection, transduction, electroporation, transfection, and transformation. Exemplary methods include DEAE-dextran mediated transfection, transient transfection, microinjection, cationic lipid-mediated transfection, scrape loading and ballistic introduction. Methods for the introduction of exogenous DNA sequences into host cells are discussed in Felgner, P. *et al.*, U.S. Patent No. 5,580,859. The nucleic acid molecules and/or vectors of the invention may be introduced alone or in conjunction with other nucleic acid molecules and/or vectors and/or proteins, peptides or RNAs. Alternatively, the nucleic acid molecules and/or expression systems of the invention may be introduced into host cells as a precipitate, such as a calcium phosphate precipitate, or in a complex with a lipid. Electroporation also may be used to introduce the nucleic acid molecules and/or expression systems of the invention into a host. Likewise, such molecules may be introduced into chemically competent cells such as *E. coli*. If the vector is a virus, it may be packaged in vitro or introduced into a packaging cell and the packaged virus may be transduced into cells. Hence, a wide variety of techniques suitable for introducing the nucleic acid molecules and/or vectors of the invention into host cells are well known and routine to those of skill in the art. Such techniques are reviewed at length, for example, in Sambrook, J., et al., *Molecular Cloning, a Laboratory Manual*, 2nd Ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 16.30-16.55 (1989), Ausubel, F. *et al.*, eds., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John H. Wiley & Sons, Inc. (1997), Chapter 16),

Watson, J.D., *et al.*, Recombinant DNA, 2nd Ed., New York: W.H. Freeman and Co., pp. 213-234 (1992), and Winnacker, E.-L., From Genes to Clones, New York: VCH Publishers (1987), which are illustrative of the many laboratory manuals that detail these techniques and which are incorporated by reference herein in their entireties for their relevant disclosures.

**[00158]** The invention is also directed to recombinant host cells comprising one or more nucleic acid molecules or expression systems of the invention. Also included within the present invention are *in vitro* cell cultures that comprise such recombinant host cells. Methods for producing recombinant host cells and cell cultures comprising the same are well-known in the art.

#### PRODUCTION OF POLYPEPTIDES AND RNA MOLECULES

**[00159]** The nucleic acid molecules, expression systems and recombinant host cells of the invention may be used for the production of proteins, polypeptides and RNA molecules, *e.g.*, untranslated RNA molecules. The methods of the invention may comprise, *e.g.*, introducing one or more nucleic acid molecules or expression systems of the present invention into host cells to produce recombinant host cells, culturing the recombinant host cells under conditions suitable for expression of the polypeptide or untranslated RNA molecule, and recovering the polypeptide or untranslated RNA molecule.

**[00160]** The invention also provides methods for regulating the expression of a polypeptide or an untranslated RNA molecule. The methods of the invention may comprise, *e.g.*, (a) introducing one or more nucleic acid molecules or expression systems of the invention into a host cell to produce a recombinant host cell, wherein said nucleic acid molecules or expression systems comprise a first open reading frame having a nucleotide sequence encoding a temperature-sensitive RNA-dependent RNA polymerase, (b) growing said recombinant host cell under suitable culture conditions, and (c) changing the temperature of the recombinant host cell culture from: (i) a permissive temperature to a restrictive temperature, or (ii) a restrictive temperature to a permissive temperature.

[00161] The present invention also provides methods for producing polypeptides and RNA molecules, said methods comprising introducing one or more nucleic acid molecules or expression systems of the invention into recombinant host cells, expanding the cells at a non-permissive temperature and incubating these cells at a permissive temperature. In a related aspect, the invention provides purified polypeptides and RNA molecules produced according to the methods of the present invention.

[00162] Depending on the molecule, which is expressed, it may be obtained either from the culture supernatant or by lysing the recombinant host cells.

[00163] Polypeptides produced using the nucleic acid molecules and expression systems of the invention can be recovered and purified from recombinant cell cultures by methods known in the art including ammonium sulfate precipitation, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and high performance liquid chromatography. Methods for purifying proteins are described in numerous sources (*see, e.g.*, Celis, J., ed., CELL BIOLOGY, Academic Press, 2<sup>nd</sup> edition, (1998)).

[00164] Untranslated RNA molecules produced using the nucleic acid molecules and expression systems of the invention can be recovered and purified from recombinant cell cultures by methods known in the art (*see, e.g.*, Celis, J., ed., CELL BIOLOGY, Academic Press, 2<sup>nd</sup> edition, (1998)). Methods for recovering and/or purifying RNA molecules include phenol/chloroform extraction, digestion with DNases followed by precipitation of the undigested RNA molecules, and column chromatography (*e.g.*, oligo dT column chromatography). Further, RNA molecules can be separated from other cellular material using the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski and Sacchi, *Anal. Biochem.* 162:156-159 (1987).

[00165] The overall cell culture process employing nucleic acid molecules and expression systems of the invention for the production of polypeptides and/or untranslated RNA molecules can be implemented in a variety of bioreactor

configurations (*e.g.*, stirred-tank, perfused, membrane enclosed, encapsulated cell, fluidized bed, and air-lift reactors) and scales (from laboratory T-flasks to thousands of liters), chosen to accommodate the requirements of the host cell line utilized (*e.g.*, anchorage dependency, O<sub>2</sub> concentrations), to maximize the production of expression product, and to facilitate subsequent recovery and purification of expression product.

[00166] The invention is also directed to the production of polypeptides or RNA molecules of interest using mammalian cells grown in serum-free or protein-free culture media. For example, by long-term culture under conditions restricting serum access or selecting for suspension growth, CHO cell lines are selected which are able to grow in serum-free medium and/or in suspension (Zang. M. *et al.*, *Bio/Technology* 13:389 (1995)).

[00167] Further, a number of different bioprocess parameters can be varied in order to alter the glycosylation pattern of polypeptide products produced by the recombinant host cells of the invention. These factors include medium composition, pH, oxygen concentration, lack or presence of agitation, and, for the case of anchorage-dependent cells, the surface provided. Thus, the glycosylation pattern of glycoproteins may be altered by choosing the host cell in which these proteins are expressed in and the conditions under which the recombinant host cells are grown.

## PHARMACEUTICAL COMPOSITIONS

[00168] The invention further provides pharmaceutical compositions comprising nucleic acid molecules and/or expression systems and/or recombinant host cells of the invention. The pharmaceutical compositions of the invention may comprise nucleic acid molecules and/or expression systems and/or recombinant host cells of the invention in combination (*e.g.*, in solution) with a physiologically acceptable carrier and in a therapeutically effective amount. The administration of these pharmaceutical compositions may, for example, result in expression of a polypeptide in tissues of an animal which is immunogenic and intended to function as a vaccination. Similarly,

the nucleic acid molecules and/or expression systems and/or recombinant host cells of the invention may carry sequences that encode polypeptides or RNA molecules required for the treatment of an active affliction. The administration of a pharmaceutical composition of the invention will thus be intended to have a therapeutic effect in these instances.

[00169] The nucleic acid molecules and/or expression systems and/or recombinant host cells of the invention will normally be administered to an individual in a pharmacologically acceptable carrier. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient individual. Further, the composition of the invention will be administered in a "therapeutically effective amount" (e.g., an amount that produces a desired physiological effect).

[00170] As would be understood by one of ordinary skill in the art, when the nucleic acid molecules and/or expression systems and/or recombinant host cells of the invention are administered to an individual, they may be in a composition which contains salts, buffers, adjuvants, or other substances which are desirable for improving the efficacy of the composition. Examples of materials suitable for use in preparing pharmaceutical compositions are provided in numerous sources including REMINGTON'S PHARMACEUTICAL SCIENCES (Osol, A, ed., Mack Publishing Co., (1980)).

[00171] The pharmaceutical compositions of the present invention can be administered by various art known means but will normally be administered by injection, infusion or other suitable physical methods. The compositions may alternatively be administered intramuscularly, intravenously, or subcutaneously. Components of compositions for administration include sterile aqueous (e.g., physiological saline) or non-aqueous solutions and suspensions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Carriers or occlusive dressings can be used to increase skin permeability and enhance antigen absorption.

[00172] When recombinant host cells are administered to an individual, the number of cells or nucleic acid molecules required to provide a therapeutically

effective amount will vary with such factors as the individual's condition, the polypeptides or RNA molecules intended to be expressed, and the size of the individual.

## KITS

**[00173]** The invention also provides kits comprising the isolated nucleic acid molecules, expression systems and/or recombinant host cells of the invention. The kits of the invention may optionally comprise one or more additional components selected from the group consisting of one or more containers (*e.g.*, boxes, vials, tubes, jars ampules, etc.) one or more vectors, one or more nucleotides, one or more primers, one or more polypeptides having polymerase activity, one or more host cells (*e.g.*, host cells that may be competent for uptake of nucleic acid molecules), and one or more buffers.

[00174] It will be readily apparent to one of ordinary skill in the relevant arts that other suitable modifications and adaptations to the methods and applications described herein are obvious and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

### EXAMPLE 1

## CONSTRUCTION OF THE pCytTs VECTOR SYSTEM

**[00175]** Manipulations and sequencing of DNA were carried out by standard procedures. The mutations in nsP2 were introduced by PCR using the following oligonucleotides:

**[00176] oligo-nsp2 1: 5'-AACATTGAAATCGATATTACAGGGG (SEQ ID NO:3),**

[00177] oligo-nsp2 2: 5'-CGGGTTATGGTTCGACCGGGC (SEQ ID NO:4),

**[00178]** oligo-nsp2 3:5'-GTGCCCTCCCCTGAGTTTAAACAATTCAGGGC  
CGAACGCG (SEQ ID NO:5), and

**[00179] oligo-nsp2 4: 5'-GAATTGTTTAAACTCAGGAGGCACCCTCGTGG**  
(SEQ ID NO:6).

**[00180]** The single restriction sites used for first analysis and subsequent cloning (*Dra*I, *Cla*I and *Sal*I) are underlined. PCR reactions were performed using either oligo-nsp2 1 (SEQ ID NO:3) and oligo-nsp2 3 (SEQ ID NO:5) or oligo-nsp2 2 (SEQ ID NO:4) and oligo-nsp2 4 (SEQ ID NO:6). 100 pmol of each oligo was used and 5 ng of the template DNA (pSinRep5; Xiong, C. *et al.*, *Science* 243:1188-1191 (1989)) was used in the 100  $\mu$ l reaction mixture, containing 4 units of Taq or Pwo polymerase, 0.1 mM dNTPs and 1.5 mM MgSO<sub>4</sub>. All DNA concentrations were determined photometrically using the GeneQuant apparatus (Pharmacia Biotech Inc., 800 Centennial Ave., Piscataway, NJ. 08854). The polymerase was added directly before starting the PCR reaction (starting point was 95°C). The temperature cycles were as follows: 95°C for 2 minutes, followed by 5 cycles of 95°C (45 seconds), 58°C (30 seconds), 72°C (90 seconds) and followed by 25 cycles of 95°C (45 seconds), 68°C (30 seconds), 72°C (90 seconds).

**[00181]** The two PCR fragment were purified using Qia spin PCR kit (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311) and finally digested in an appropriate buffer using 20 units of *SalI* and *DraI*, respectively 20 units of *Clal* and *DraI*. The digestion was performed for 12 hours at 37°C. The DNA fragments were gel-purified (Gene-Clean; Bio 101 Inc., 1070 Joshua Way, Vista, CA, 92083, USA) and finally ligated into *Clal/SalI* digested and gel-purified SinRep5 vector (Xiong, C. *et al.*, *Science* 243:1188-1191 (1989). The correct sequence of the obtained vector was checked by DNA sequencing of the whole nsP2 gene.

**[00182]** The mutations in nsP4 were also introduced by PCR using the following oligonucleotides:

**[00183]** oligo-nsp4 1: 5'-GGTAGACGAGACAGTCGCATGCCTGGATAC  
(SEQ ID NO:7),



[00184] oligo-nsp4 2: 5'-GTATCCAGGCATGCGACTGTCTCGTCTACC  
(SEQ ID NO:8),

[00185] oligo-nsp4 3: 5'-CAGACCGGTTAACGCCATAGCG TCG (SEQ ID  
NO:9), and

[00186] oligo-nsp4 4: 5'-CTCTATTACTAGTATGGACAGTTGG (SEQ ID  
NO:10),

[00187] The singular restriction sites used for the first analysis and the final cloning step (*Sph*I, *Hpa*I and *Spe*I) are underlined. Two PCR reactions were carried out as described above using either oligo-nsp4 1 (SEQ ID NO:7) and oligo-nsp4 3 (SEQ ID NO:9) or oligo-nsp4 2 (SEQ ID NO:8) and oligo-nsp4 4 (SEQ ID NO:10).

[00188] Both PCR products were gel-purified and then used in assembly PCR to amplify the whole nsP4 gene. For the assembly PCR, 50 pmol of the outer primers (3 and 4) and about 10 ng of each PCR fragment was used. The reaction volume was 100 µl, containing 4 units of Taq or Pwo polymerase, 0.1 mM dNTPs and 1.5 mM MgSO<sub>4</sub>. The PCR conditions were as followed:

[00189] 95°C for 2 minutes, followed by 5 cycles of 92°C (45 seconds), 58°C (30 seconds), 72°C (120 seconds) and followed by 25 cycles of 92°C (45 seconds), 64°C (30 seconds), 72°C (120 seconds).

[00190] The obtained PCR fragment was purified as described above and the eluate was digested with 20 units of *Spe*I and *Hpa*I in an appropriate buffer. The fragment was gel-purified and ligated into gel-purified *Spe*I/*Hpa*I restricted SinRep5 vector. The correct sequence of the obtained vector was checked by DNA sequencing.

[00191] Overnight digestion of SinRep5-nsP4mut and SinRep5-nsp2mut with *Spe*I/*Hpa*I and gel purification of the nsp4 fragment and sinRep-nsp2mut vector. The nsp4mut fragment was ligated into the SinRep5-nsp2mut vector. The final step was cloning the nsp gene into the 987/SinRep5 vector (Bredenbeek, P. *et al.*, *J. Virol.* 67:6439-6446 (1993)) using *Cla*I and *Hpa*I as restriction endonucleases, the resulting vector was named pCytTs (SEQ ID NO:1, Fig. 1).

[00192] Vector pCytTS2.1 (SEQ ID NO:2, Fig. 2) contains a different multiple cloning site than vector pCytTS, the RSV promoter is replaced by the CMV promoter, an SV40 intron sequence is included in the replicon and a puromycin resistance marker under control of the SV40 promoter is integrated on the plasmid. These elements can be cloned into pCytTS (SEQ ID NO: 1) in the following way:

[00193] Oligonucleotides Cyt-Link-FOR 5'- CTAGATTAATTAAC TCGAG GCGCGCCG-3' (SEQ ID NO: 11) and Cyt-Link-Rev 5'- GGCCCGGCGCGCCTCGAGTTAATTAAT-3' (SEQ ID NO:12) are hybridized and ligated into *XbaI/BspI*20I digested vector pCytTS. This step introduces a different multiple cloning site into pCytTS (Intermediate).

[00194] The CMV promoter is introduced by assembled PCR using the following oligonucleotides:

[00195] 5'CMV: 5'-ATAAGAATGCCGGCGGATCCGGCCATTAGC-3' (SEQ ID No:13)

[00196] 3'SinCMV: 5'-CCGTCAATACGGTTCCTAAACGAGCTCTG GTTATATAGACC-3' (SEQ ID No: 14)

[00197] 5'CMV Sin: 5'-GCTCGTTTAGTGAACCGTATTGACGGCGTA GTACACAC-3' (SEQ ID No: 15)

[00198] 3'Sin: 5'- ACGTCGGCCTCAATTTCGCG-3' (SEQ ID No: 16)

[00199] The singular restriction site used for the final cloning step (*NgoAIV*) is underlined. Two PCR reactions are carried out as described above using either oligonucleotide 5'CMV 1 (SEQ ID NO:13) and 3'SinCMV (SEQ ID NO:14) or 5'CMV Sin (SEQ ID NO:15) and 3'Sin (SEQ ID NO:16). As template DNA vector pLNCX (Miller et al., 1989, *Biotechniques*, 7, 980-982) and vector pCytTS can be used, respectively. Both PCR products are gel-purified and then used in assembly PCR. For the assembly PCR the outer primers (5'CMV and 3'Sin) are used, resulting in an 3089 bp size PCR fragment. The obtained PCR fragment is purified as described above and the eluate is digested with *NgoAIV* and *Eco47III* in an appropriate buffer. The fragment is gel-purified and ligated into gel-purified *NgoAIV/Eco47III*

restricted IntermediateI. This cloning step replaces the RSV promoter in pCytTS with the CMV promoter (IntermediateII).

**[00200]** The SV40 intron is amplified by PCR from vector pcDNA1.1amp (Invitrogen) using the following oligonucleotides:

**[00201] SVIntron-FOR: 5'-GCGCGCGGGCCAGAGGATCTTTGTGA**  
**AGG-3' (SEQ ID NO:17)**

**[00202] SVIntron-REV: 5'-GCGCGCGCGGCCGCTACATCAAATATTT**  
**TTCC-3' (SEQ ID NO:18)**

**[00203]** The PCR conditions can be chosen as followed: 94°C for 2 minutes, followed by 30 cycles of 94°C (45 seconds), 52°C (30 seconds), 72°C (60 seconds) and followed by a 7 min elongation step at 72°C. The obtained PCR fragment is purified using the Qia spin PCR kit (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). Finally the purified PCR fragment is digested in the corresponding buffer with *Bsp120I*/*NorI* and ligated into *Bsp120I* digested IntermediateII. The orientation of the insert is controlled by restriction enzyme analysis (IntermediateIII). The SV40 intron is thus introduced behind the multiple cloning site.

**[00204]** To introduce the puromycin resistance marker, vector pPUR (Clontech Laboratories, Inc., Cat. No. 6156-1) is modified first. Vector pPUR is digested with *Xba*I and the linearized vector is treated with Klenow polymerase in order to get blunt ends. Religation of the vector destroys the *Xba*I site. The resulting vector serves as DNA template in the PCR amplification of the puromycin resistance marker. The following oligonucleotides are used:

**[00205]** 5'SV40PUR: 5'- ACGTACGCGTGCGGCCGCGTTAGGGTGTGG  
AAAGTCCCC-3' (SEQ ID NO:19)

**[00206] 3'SV40PUR: 5'- ACGTACGCGTTGGACAAACCACA ACTAG**  
**AATGC-3' (SEQ ID NO:20)**

**[00207]** The singular restriction site used for the cloning step (*Mlu*I) is underlined. PCR conditions can be chosen as described before with an annealing temperature at 45°C. The resulting PCR fragment is purified as described above, digested with *Mlu*I and ligated into *Mlu*I digested

Intermediate III. The resulting vector pCytTs2.1 (SEQ ID NO:2, Fig. 2) contains the puromycin resistance marker under control of the SV40 promoter.

## EXAMPLE 2

### COMBINATION OF THE TEMPERATURE SENSITIVE pCytTs SYSTEM WITH THE SELF REPLICATING EBNA SYSTEM

[00208] This system was generated in order to rapidly generate cell populations, which inducibly express a gene of interest. In this system, the tightly regulated pCytTs system is combined with the episomally replicating EBV system.

A. Construction of various pCytTs constructs containing either the EBNA origin of replication (OriP) alone or in conjunction with the replication initiation factor EBNA1.

[00209] In order to be able to subclone either the origin of replication of Epstein-Barr virus alone or together with the replication initiation factor EBNA1 we first subcloned this two cassettes into a shuttle vector from which they can easily be transferred into our pCytTs system with or without various inserts. Combinatorial constructs with pCytTs mentioned in the Examples such as pCytTs-OriP, pCytTs-SEAP-OPE, pCytTs-SEAP or the like generally contain a CMV promoter. The sequence of the pCytTs with the CMV promoter is given in SEQ ID NO: 2 (pCytTs2.1; see Example 1). For sake of simplicity, however, the mentioned combinatorial constructs are abbreviated by simply stating pCytTs. The Epstein-Barr virus origin of replication OriP was amplified by PCR from pCEP4 (Invitrogen) using primers PH 51 and PH 52. The forward primer (PH 51) corresponds to nucleotides 261-280 of pCep4 and the reverse primer (PH 52) to nucleotides 2221-2239 of pCep4. PCR was performed using pfx polymerase (BRL) according to manufacturer's recommendations. The resulting PCR product (nucleotides 261-2239 of pCep4) was then cloned into pGEMTeasy vector (Invitrogen), which had

previously been digested with Eco RI and treated with Klenow polymerase in order to get blunt ends. The resulting plasmid was termed pGemT-OriP. In order to get a subclonable cassette containing OriP and EBNA1 gene, pCep4 was digested with Dra III and Cla I. The resulting 4651 bp fragment was purified and blunt ends were generated by T4 DNA polymerase treatment. This fragment was then subcloned into pGEMTeasy vector (Invitrogen), which had previously been digested with Eco RI and treated with Klenow polymerase in order to get blunt ends. The resulting construct was termed pGemT-OPE. From these plasmids either the origin of replication alone (pGemT-OriP) or together with the replication initiation factor (pGemT-OPE) could be excised by Not I digestion and transferred into Not I sites of the various CytTs vectors (e.g. pCytTs vectors containing the RSV or CMV promoter (SEQ ID NO: 1 or SEQ ID NO: 2) which resulted in the vectors pCytTs-OriP (SEQ ID NO: 21), and pCytTs-OPE (SEQ ID NO: 22). pCytTs-IFN $\beta$ , pCytTs-EGFP, pCytTs-SEAP and pCytTs-cEPO vectors were linearized with Not I and the Not I fragment of pGemT-OPE was introduced by ligation leading to following constructs pCytTs-IFN $\beta$ -OPE pCytTs-EGFP-OPE (Fig. 3), pCytTs-SEAP-OPE and pCytTs-EPO-OPE. Similarly pCytTs-EPO-OriP and pCytTs-EGFP-OriP were generated by cloning the Not I fragment from pGemT-OriP into the Not I site of pCytTs-cEPO and pCytTs-EGFP respectively.

[00210] Initial experiments were performed with a construct in which a slightly truncated form of OriP had been subcloned into pCytTs-EGFP. pCep4 (Invitrogen) was digested with Dra III and Nsi I and the 1923 nucleotide fragment (position 200-2124 from pCep4) was then treated with T4 polymerase in order to get blunt ends. The blunted fragment was then subcloned into pCytTs-EGFP, which had previously been digested with Not I and blunted with T4 DNA polymerase. The resulting plasmid was termed pCytTs-EGFP-OriPd.

#### B. Analysis of pCytTs-EGFP-OriPd in 293 EBNA cells

[00211] We first analyzed whether the introduction of Epstein-Barr virus derived origin of replication (OriP) into the pCytTs system would improve the inducible expression of a gene of interest on selected cell populations. We therefore transfected 293-EBNA cells (Invitrogen) with pCytTs-EGFP-OriPd. These cells constitutively express Epstein-Barr virus replication initiation factor EBNA1 and support OriP mediated DNA replication. 293-EBNA cells were transfected with pCytTs-EGFP-OriPd using Lipofectamine 2000 (BRL) according to the manufacturer's recommendation. 36h after transfection cells were split 1 to 3 in presence of 1µg/ml puromycin. After 1 to 2 days all non-transfected cells had detached and the resistant cells were further expanded under puromycin selection for another week. In order to assess the expression from the new vector, resistant cell populations were passed to two plates of which one was kept at 37°C (uninduced) and one was shifted to 29°C for induction of EGFP expression. 24h after induction cells were harvested by trypsinisation and analysed by flow cytometry (Fig. 4). Upon induction cell populations carrying pCytTs-EGFP-OriPd were to 80% EGFP positive and displayed strong EGFP expression. In contrast cells, which had not been induced showed only weak EGFP expression and only about 8% of the cells, were green. Considering that stable cell populations containing CytTs-EGFP would only yield about 2-5% EGFP positive cells upon induction (data not shown) these results suggest that the OriP, which was introduced in the new vector, is functional and that it significantly improves the number of cells in a cell population, which expresses the gene of interest.

C. Comparison of pCytTs-EGFP, pCytTs-EGFP-OriP and pCytTs-EGFP-OPE in BHK cells

[00212] The following constructs, pCytTs-EGFP, pCytTs-EGFP-OriP and pCytTs-EGFP-OPE, were next examined in BHK (Baby hamster kidney) cells which are known to show tight regulation for the CytTs system. BHK cells

[00213] We further analyzed constructs which contain both OriP and the replication initiation factor EBNA1 (pCytTs-IFN $\beta$ -OPE) with other genes of interest in BHK cells. These constructs do not depend on cell lines which have an integrated copy of the EBNA1 gene. As a gene of interest we choose IFN $\beta$ , which had already been expressed with the conventional CytTs system. To test whether our new vector system with its episomal replication could result in high, inducible levels of expression on selected cell populations we transfected BHK cells with pCytTs-IFN $\beta$ -OPE (*see* Example 2A) using Lipofectamine 2000 (BRL) according to the manufacturer's recommendation. 36h after transfection cells were split 1 to 3 in the presence of 5 $\mu$ g/ml puromycin. After 1 to 2 days all non-transfected cells had detached and the resistant cells were further expanded under puromycin selection for another 1 -

2 weeks. In order to assess the expression from the new vector, resistant cell populations were passed to two plates one of which was kept at 37°C (uninduced) and one was shifted to 29°C for induction of IFN $\beta$  expression. Tissue culture supernatants were harvested from the uninduced control plate and 3, 4, 7 and 9 day after induction from the plate that had been shifted to 29°C. The IFN $\beta$  production at the different time points was then quantitated by ELISA (Kit of Fujirebio Inc., Tokyo, Japan). The induced cell populations reached a maximal level of 1,000,000 IU/ml after 7-9 days of induction and no IFN production could be detected in cells kept at non-permissive temperature (Figs. 6A-6B). These results show that our new vector system combining the CytTs system with the episomally replicating EBNA system leads to high production levels in BHK cells on selected populations. The expression levels are at least as high as the ones obtained from a stable cell clone, which resulted from 4 subcloning rounds (data not shown). In addition the new system is still tightly regulated as demonstrated by up to 300,000 fold induction observed upon shifting the cells to 29°C.

**[00214]** Besides IFN $\beta$  production we also analysed SEAP production with our new vector system in BHK cells. Selected cell populations were generated as described under C using construct pCytTs-SEAP-OPE. After two weeks of selection and propagation the resistant cells were passed into two plates of which one was kept at 37°C and the other was shifted to 29°C for induction. Samples were taken from the uninduced cell populations and from the induced cell population 3, 4, 7 and 9 days after induction. SEAP concentrations were measured in these samples using a kinetic ELISA for SEAP activity. The results are shown in Fig. 7.

**E. Comparison of the EPO production derived from pCytTs-OriP-EPO or with pCytTs-OPE-EPO in 293 EBNA cells**

**[00215]** We further investigated whether an additional copy of the EBNA-1 gene on the same plasmid leads to higher expression of the gene of interest in



293-EBNA cells. 293-EBNA cells were transfected either with pCytTs-OriP-EPO or with pCytTS-OPE-EPO (see Example 2A) using Lipofectamin 2000 (BRL) according to the manufacturer's recommendation. One day after transfection the cells were split 1 to 2.5 in the presence of 0.8  $\mu\text{g}/\text{mL}$  puromycin. A stable cell pool was obtained by further keeping these cells under selective pressure for one week. Resistant cell populations were then passed to two plates. Once the cells had attached, one plate was kept at 37°C and one plate was shifted to 29°C for induction. Cell culture supernatants were collected from both plates 4 and 8 days after induction. The erythropoietin (EPO) levels in the cell culture supernatants were then determined with an ELISA kit (R&D Systems, Minneapolis, USA). As shown in Fig. 8 the additional copy of the EBNA1 gene on the plasmid leads to a 15 to 20 fold increase in the EPO production compared to the plasmid which only contains the OriP. In conclusion, this experiment shows that, even in cells which contain an integrated copy, the expression of the gene of interest can be increased by an additional copy on the plasmid. These results suggest that the EBNA1 dosage may be a limiting factor for efficient replication to occur. In these experiments the EBNA1 dosage was increased by a copy on the OriP-containing plasmid. Similar results may be obtained if the EBNA1 dosage is increased in the cells either by the integration of further copies or if the additional copies would be provided on a separate plasmid in *trans*.

### EXAMPLE 3

## CONSTRUCTION OF VECTORS OF THE pCytTs SYSTEM CONTAINING THE GLUTAMINE SYNTHETASE AS SELECTABLE MARKER

**[00216]** The glutamine synthetase gene is cloned either from hamster cells (CHO-K1, Chinese hamster ovary, ATCC, Cat. No. CCL-61; BHK21, Hamster Syrian kidney cells, DSMZ, Cat. No. ACC 61) or human cells (HEK 293 cells (ATCC, Cat. No. CRL-1573; Hela ATCC, Cat. No. CCL-2; Raji

cells, ATCC, Cat. No. CCL-86; 293 EBNA, Invitrogen, Cat. No. R62007; and 143B cells, ATCC, Cat. No. CRL-8303) by the method of reverse transcription. Total RNA or cytoplasmic RNA is isolated from the cells using the RNeasy Kit (Qiagen, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311) according to the manufactures recommendation. Enrichment of poly(A)<sup>+</sup> RNA from total RNA can be done by using the Oligotex mRNA Kit (Qiagen, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). The resulting RNA can be treated with DNaseI to remove residual traces of DNA. cDNA synthesis is performed in the first step using the ThermoScript RT-PCR System (Invitrogen, Cat. No. 11146-024) with either total RNA or poly(A)<sup>+</sup>-selected RNA primed with oligo(dT) or random primers according to the manufactures recommendation. Treatment of cDNA with RNase H to remove the complementary RNA prior to PCR is optional. In the second step, PCR is performed using primers specific for the hamster or human glutamine synthetase. To amplify the cDNA of hamster and human glutamine synthetase the following oligonucleotides can be used:

[00217] hamGSfor: 5'- GTCATGAAGCTTGCCACCGCTCAGAGC  
ACCTTC-3' (SEQ ID NO: 23)

[00218] ham GSrev: 5'- GACTTCTAGACTGGGGCGGGGTGGGATGAAC-  
3' (SEQ ID NO: 24)

[00219] hGSfor: 5'- GTCATGAAGCTTCTCGGCGACCAGAACACCTTC-3'  
(SEQ ID NO: 25)

[00220] hGSrev: 5'- GACTTCTAGACTGGGGCGGGGTGGGATGAAC  
(SEQ ID NO: 26)

[00221] Restriction sites used for cloning are underlined (*Hind*III, *Xba*I). The resulting PCR fragment is purified using Qia spin PCR kit (Qiagen, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311).

[00222] To replace the puromycin resistance marker with the glutamine synthetase cassette in vector pCytTS2.1, the glutamine synthetase cassette is isolated from *Mlu*I and *Bam*HI digested vector pGL3-GS. The fragment is treated with Klenow polymerase in order to get blunt ends and ligated into *Mlu*I digested vector pCytTS2.1 treated with Klenow polymerase.

[00223] To combine the resulting pCytTS2.1gs vector with the self replicating EBNA system either the origin of replication alone (pGEMT-OriP) or together with the replication initiation factor (pGemT-OriP) can be excised by *NotI* digestion and transferred into the *NotI* site of pCytTS2.1gs.

[00224] Vectors of the present invention containing the glutamine synthetase gene as selection marker are transfected in different cell lines as described above (EXAMPLE 2). Mammalian cell lines that do not express GS (glutamine synthetase) cannot survive without added glutamine in the medium. For these cell lines, a transfected GS gene can function as a selectable marker by permitting growth in a glutamine-free medium (e.g. mouse myeloma NS0 cell line). For other cell lines (e.g. CHO-K1) producing endogenous GS, methionine sulfoximine (MSX) has to be added as selective agent to the glutamine-free medium.

#### EXAMPLE 4

##### STABILITY OF EXPRESSION IN DIFFERENT CELL LINES USING THE pCytTS-EBNA SYSTEM

[00225] Vectors of the pCytTS-EBNA system containing either the EBNA origin of replication (OriP) alone or in conjunction with the replication initiation factor EBNA1 were transfected in different cell lines.

[00226] 293 EBNA cells were transfected with vector pCytTS-EPO-OriP using Lipofectamine 2000 (BRL) according to the manufacturer's recommendation. One day after transfection cells were split 1 to 4 in the presence of 0.8 µg/ml puromycin. Several days after selection a stable cell pool was obtained which was further expanded. To assess inducible EPO expression, resistant cell populations were trypsinized and two T25 flask were seeded with three million cells each. Once the cells had attached one T25 flask was kept at 37 °C (uninduced) and one was shifted to 29 °C for induction of replicon replication and EPO expression. Cell culture supernatants were collected from both plates seven days after induction. The erythropoietin (EPO) level in the cell culture

-62-

supernatants was determined with an ELISA kit (R&D Systems, Minneapolis, USA). A backup plate of the resistant cell pool was kept in culture and passaged about 1 to 10 times a week. To determine stability of EPO expression over several weeks, cells from the backup plate were seeded at intervals of one or two weeks again on T25 flask as described above. Cells were induced for 7 days and EPO expression was determined by ELISA (Fig. 9).

[00227] These results show that stable cell pools containing a vector system combining the pCytTS system with the episomally replicating EBNA system allow expression of EPO over several weeks without loss of productivity. EPO values in the range of 3-5 mg/l are obtained. Further on the pCytTS system is tightly regulated as at 37°C hardly no EPO production can be detected.

[00228] Having now fully described the present invention in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious to one of ordinary skill in the art that the same can be performed by modifying or changing the invention within a wide and equivalent range of conditions, formulations and other parameters without affecting the scope of the invention or any specific embodiment thereof, and that such modifications or changes are intended to be encompassed within the scope of the appended claims.

[00229] All publications, patents and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains, and are herein incorporated by reference to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. A nucleic acid molecule comprising:
  - (A) a first polynucleotide element which encodes an RNA molecule, said RNA molecule comprising:
    - (a) at least one *cis*-acting sequence element;
    - (b) a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and
    - (c) at least one second nucleotide sequence selected from the group consisting of:
      - (i) a second open reading frame encoding a polypeptide;
      - (ii) a nucleotide sequence complementary to all or a part of the second open reading frame of (i); and
      - (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof;wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase;
  - (B) a second polynucleotide element comprising an origin of replication; and
  - (C) a third polynucleotide element encoding a replication initiation factor capable of recognizing said origin of replication.
2. The nucleic acid molecule of claim 1, wherein said second open reading frame is in a translatable format after one or more RNA-dependent RNA replication events.
3. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is selected from the group consisting of: (a) a temperature-sensitive RNA-dependent RNA polymerase; (b) a non-

cytopathic RNA-dependent RNA polymerase; and (c) a temperature-sensitive, non-cytopathic RNA-dependent RNA polymerase.

4. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a prokaryotic organism or a eukaryotic organism.
5. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a virus.
6. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a DNA virus.
7. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a DNA virus that allows for episomal replication.
8. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a DNA virus selected from the group consisting of Papillomavirus, Polyomavirus, Adenovirus, and Hepadnaviruses.
9. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a Herpesvirus.
10. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from Epstein-Barr virus (EBV).
11. The nucleic acid molecule of claim 10, wherein said origin of replication is *oriP*.

-65-

12. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from an eukaryotic organism selected from the group consisting of yeast, and insects.
13. The nucleic acid molecule of claim 1, wherein said origin of replication is derived from a mammal.
14. The nucleic acid molecule of claim 1, wherein said replication initiation factor is capable of operating as a plasmid maintenance factor.
15. The nucleic acid molecule of claim 1, wherein said replication initiation factor is derived from a prokaryotic organism or a eukaryotic organism.
16. The nucleic acid molecule of claim 1, wherein said replication initiation factor is derived from a virus.
17. The nucleic acid molecule of claim 1, wherein said replication initiation factor is derived from a DNA virus.
18. The nucleic acid molecule of claim 1, wherein said replication initiation factor is derived from a DNA virus selected from the group consisting of Herpesvirus, Papillomavirus, Polyomavirus, Adenovirus, and Hepadnavirus.
19. The nucleic acid molecule of claim 1, wherein said replication initiation factor is derived from a Herpesvirus.
20. The nucleic acid molecule of claim 1, wherein said replication initiation factor is derived from EBV.

21. The nucleic acid molecule of claim 20, wherein said replication initiation factor is the EBNA-1 protein.
22. The nucleic acid molecule of claim 1, wherein said replication initiation factor and said origin of replication are derived from the same organism or the same virus.
23. The nucleic acid molecule of claim 1, wherein said replication initiation factor and said origin of replication are derived from different organisms or viruses.
24. The nucleic acid molecule of claim 1 further comprising a fourth polynucleotide element, wherein said fourth polynucleotide element comprises a selection marker.
25. The nucleic acid molecule of claim 24, wherein said selection marker confers resistance to Puromycin.
26. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is of viral origin.
27. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is of alphaviral origin.
28. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is derived from a Sindbis virus.
29. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is derived from a Semliki Forest virus.
30. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is derived from an Aura virus.



31. The nucleic acid molecule of claim 1, wherein said RNA-dependent RNA polymerase is derived from a virus selected from the group consisting of Bebaru virus, Cabassou virus, Chikungunya virus, Easter equine encephalomyelitis virus, Fort morgan virus, Getah virus, Kyzylagach virus, Mayoaro virus, Middleburg virus, Mucambo virus, Ndumu virus, Pixuna virus, Tonate virus, Trinita virus, Una virus, Western equine encephalomyelitis virus, Whataroa virus, Venezuelan equine encephalomyelitis virus (VEE), and Ross River virus.
32. The nucleic acid molecule of claim 3, wherein said temperature-sensitive RNA-dependent RNA polymerase has replicase activity at temperatures below 34°C which is at least five fold greater than the replicase activity exhibited at 34°C or above.
33. The nucleic acid molecule of claim 3, wherein said temperature-sensitive RNA-dependent RNA polymerase has replicase activity at 34°C which is at least five fold lower than the replicase activity exhibited at 29°C.
34. The nucleic acid molecule of claim 1, wherein said second open reading frame encodes a cytokine, a lymphokine, a tumor necrosis factor, an interferon, a toxic polypeptide, a prodrug or a converting enzyme.
35. The nucleic acid molecule of claim 1, wherein said second open reading frame encodes human  $\beta$ -interferon.
36. The nucleic acid molecule of claim 1, wherein said second nucleotide sequence encodes an untranslated RNA molecule selected from the

group consisting of an antisense RNA molecule, a tRNA molecule, a rRNA molecule, and a ribozyme.

37. An expression system comprising:
- (A) a first polynucleotide element which encodes an RNA molecule, said RNA molecule comprising:
    - (a) at least one *cis*-acting sequence element;
    - (b) a first nucleotide sequence comprising a first open reading frame, said first open reading frame having a nucleotide sequence encoding an RNA-dependent RNA polymerase; and
    - (c) at least one second nucleotide sequence selected from the group consisting of:
      - (i) a second open reading frame encoding a polypeptide;
      - (ii) a nucleotide sequence complementary to all or a part of the second open reading frame of (i); and
      - (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof;wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase;
  - (B) a second polynucleotide element comprising an origin of replication; and
  - (C) a third polynucleotide element encoding a replication initiation factor capable of recognizing said origin of replication.
38. The expression system of claim 37, wherein said second open reading frame is in a translatable format after one or more RNA-dependent RNA replication events.

39. The expression system of claim 37, wherein said first, second and third polynucleotide elements are DNA elements.
40. The expression system of claim 37, wherein said first, second and third polynucleotide elements are each present on a separate nucleic acid molecule.
41. The expression system of claim 37, wherein said first, second and third polynucleotide elements are all present on a single nucleic acid molecule.
42. The expression system of claim 37, wherein said first and said second polynucleotide elements are present on a first nucleic acid molecule, and said third polynucleotide element is present on a second nucleic acid molecule.
43. The expression system of claim 42, wherein said second nucleic acid molecule further comprises a fourth polynucleotide element capable of promoting the replication of said second nucleic acid molecule.
44. The expression system of claim 42, wherein said second nucleic acid sequence further comprises an origin of replication.
45. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is selected from the group consisting of: (a) a temperature-sensitive RNA-dependent RNA polymerase; (b) a non-cytopathic RNA-dependent RNA polymerase; and (c) a temperature-sensitive, non-cytopathic RNA-dependent RNA polymerase.
46. The expression system of claim 37, wherein said origin of replication is derived from a prokaryotic organism or a eukaryotic organism.

-70-

47. The expression system of claim 37, wherein said origin of replication is derived from a virus.
48. The expression system of claim 37, wherein said origin of replication is derived from a DNA virus.
49. The expression system of claim 37, wherein said origin of replication is derived from a DNA virus that allows for episomal replication.
50. The expression system of claim 37, wherein said origin of replication is derived from a DNA virus selected from the group consisting of Herpesvirus, Papillomavirus, Polyomavirus, Adenovirus, and Hepadnaviruses.
51. The expression system of claim 37, wherein said origin of replication is derived from a Herpesvirus.
52. The expression system of claim 37, wherein said origin of replication is derived from Epstein-Barr virus (EBV).
53. The expression system of claim 52, wherein said origin of replication is *oriP*.
54. The expression system of claim 37, wherein said origin of replication is derived from an eukaryotic organism selected from the group consisting of yeast, insects and mammals.
55. The expression system of claim 37, wherein said origin of replication is derived from a mammal.
56. The expression system of claim 37, wherein said replication initiation factor is capable of operating as plasmid maintenance factor.

-71-

57. The expression system of claim 37, wherein said replication initiation factor is derived from a prokaryotic organism or a eukaryotic organism.
58. The expression system of claim 37, wherein said replication initiation factor is derived from a virus.
59. The expression system of claim 37, wherein said replication initiation factor is derived from a DNA virus.
60. The expression system of claim 37, wherein said replication initiation factor is derived from a DNA virus selected from the group consisting of Herpesvirus, Papillomavirus, Polyomavirus, Adenovirus, and Hepadnavirus.
61. The expression system of claim 37, wherein said replication initiation factor is derived from a Herpesvirus.
62. The expression system of claim 37, wherein said replication initiation factor is derived from EBV.
63. The expression system of claim 37, wherein said replication initiation factor is the EBNA-1 protein, or portion thereof.
64. The expression system of claim 37, wherein said replication initiation factor and said origin of replication are derived from the same organism or the same virus.
65. The expression system of claim 37, wherein said replication initiation factor and said origin of replication are derived from different organisms.

66. The expression system of claim 37, further comprising a fourth polynucleotide element, wherein said fourth polynucleotide element comprises a selection marker.
67. The expression system of claim 66, wherein said selection marker confers resistance to Puromycin.
68. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is of viral origin.
69. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is of alphaviral origin.
70. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is derived from a Sindbis virus.
71. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is derived from a Semliki Forest virus.
72. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is derived from an Aura virus.
73. The expression system of claim 37, wherein said RNA-dependent RNA polymerase is derived from a virus selected from the group consisting of Bebaru virus, Cabassou virus, Chikungunya virus, Easter equine encephalomyelitis virus, Fort morgan virus, Getah virus, Kyzylagach virus, Mayoaro virus, Middleburg virus, Mucambo virus, Ndumu virus, Pixuna virus, Tonate virus, Trinita virus, Una virus, Western equine encephalomyelitis virus, Whataroa virus, Venezuelan equine encephalomyelitis virus (VEE), and Ross River virus.

74. The expression system of claim 45, wherein said temperature-sensitive RNA-dependent RNA polymerase has replicase activity at temperatures below 34°C which is at least five fold greater than the replicase activity exhibited at 34°C or above.
75. The expression system of claim 45, wherein said temperature-sensitive RNA-dependent RNA polymerase has replicase activity at 34°C which is at least five fold lower than the replicase activity exhibited at 29°C.
76. The expression system of claim 37, wherein said second open reading frame encodes a cytokine, a lymphokine, a tumor necrosis factor, an interferon, a toxic protein, a prodrug or a converting enzyme.
77. The expression system of claim 37, wherein said second open reading frame encodes human  $\beta$ -interferon.
78. The expression system of claim 37, wherein said second nucleotide sequence encodes an untranslated RNA molecule selected from the group consisting of an antisense RNA molecule, a tRNA molecule, a rRNA molecule, and a ribozyme.
79. A method of making a recombinant host cell comprising introducing the nucleic acid molecule of claim 1 into a host cell.
80. The method of claim 79, wherein said RNA-dependent RNA polymerase is selected from the group consisting of: (a) a temperature-sensitive RNA-dependent RNA polymerase; (b) a non-cytopathic RNA-dependent RNA polymerase; and (c) a temperature-sensitive, non-cytopathic RNA-dependent RNA polymerase.





-75-

- (ii) a nucleotide sequence complementary to all or a part of the open reading frame of (i); and
- (iii) a nucleotide sequence encoding an untranslated RNA molecule or complement thereof;

wherein said second nucleotide sequence is operably linked to a promoter which is recognized by said RNA-dependent RNA polymerase;

- (B) a second polynucleotide element comprising an origin of replication; and
  - (C) a third polynucleotide element encoding a replication initiation factor capable of recognizing said origin of replication.
88. The recombinant host cell of claim 87, wherein said second open reading frame is in a translatable format after one or more RNA-dependent RNA replication events.
89. The recombinant host cell of claim 87, wherein said first, second and third polynucleotide elements are DNA elements.
90. The recombinant host cell of claim 87, wherein said third polynucleotide is stably integrated into the genome of said host cell.
91. The recombinant host cell of claim 87, wherein said first and said second polynucleotide elements are present on a first nucleic acid molecule, and said third polynucleotide element is present on a second nucleic acid molecule.
92. The recombinant host cell of claim 87, further comprising a fourth polynucleotide comprising a selection marker.
93. The recombinant host cell of claim 92, wherein said selection marker confers resistance to Puromycin.

-76-

94. The recombinant host cell of claim 91, wherein said second nucleic acid molecule further comprises a fifth polynucleotide element capable of promoting the replication of said second nucleic acid molecule.
95. The recombinant host cell of claim 91, wherein said nucleotide sequence further comprises an origin of replication.
96. The recombinant host cell of claim 87, wherein said first, second and third polynucleotide elements are provided on the same nucleic acid molecule.
97. The recombinant host cell of claim 87, wherein said RNA-dependent RNA polymerase is selected from the group consisting of: (a) a temperature-sensitive RNA-dependent RNA polymerase; (b) a non-cytopathic RNA-dependent RNA polymerase; and (c) a temperature-sensitive, non-cytopathic RNA-dependent RNA polymerase.
98. The recombinant host cell of claim 87, wherein said recombinant host cell is a mammalian cell.
99. The recombinant host cell of claim 98, wherein said mammalian cell is selected from the group consisting of a human cell, a primate cell and a rodent cell.
100. The recombinant host cell of claim 87, wherein said host cell comprises more than one copy of said third polynucleotide.
101. The recombinant host cell of claim 100, wherein at least one copy of said third polynucleotide is stably integrated into the genome of said host cell.

-77-

102. A method for producing a polypeptide or untranslated RNA molecule, said method comprising:
  - (a) introducing the nucleic acid molecule of claim 1 into a host cell to produce a recombinant host cell; and
  - (b) culturing said recombinant host cell under conditions suitable for expression of said polypeptide or untranslated RNA molecule.
103. The method of claim 102, further comprising recovering said polypeptide or untranslated RNA molecule.
104. A method for producing a polypeptide or untranslated RNA molecule, said method comprising:
  - (a) introducing the expression system of claim 37 into a host cell to produce a recombinant host cell; and
  - (b) culturing said recombinant host cell under conditions suitable for expression of said polypeptide or untranslated RNA molecule.
105. The method of claim 104, further comprising recovering said polypeptide or untranslated RNA molecule.
106. A method for regulating the expression of a polypeptide or an untranslated RNA molecule, said method comprising:
  - (a) introducing the nucleic acid molecule of claim 3 into a host cell to produce a recombinant host cell;
  - (b) growing said recombinant host cell under suitable culture conditions; and
  - (c) changing the temperature of the recombinant host cell culture from:
    - (i) a permissive temperature to a restrictive temperature, or
    - (ii) a restrictive temperature to a permissive temperature;

wherein said polypeptide or untranslated RNA molecule is encoded by said first polynucleotide.

107. The method of claim 106, wherein said polypeptide is a polypeptide that is toxic to said host cell.
108. A method for regulating the expression of a polypeptide or an untranslated RNA molecule, said method comprising:
- (a) introducing the expression system of claim 45 into a host cell to produce a recombinant host cell;
  - (b) growing said recombinant host cell under suitable culture conditions; and
  - (c) changing the temperature of the recombinant host cell culture from:
    - (i) a permissive temperature to a restrictive temperature, or
    - (ii) a restrictive temperature to a permissive temperature;

wherein said polypeptide or untranslated RNA molecule is encoded by said first polynucleotide.

109. The method of claim 108, wherein said polypeptide is a polypeptide that is toxic to said host cell.

## GENE EXPRESSION SYSTEM

### ABSTRACT OF THE DISCLOSURE

The present invention relates to compositions and methods that allow the production of polypeptides and untranslated RNA molecules in host cells. More specifically, the invention provides nucleic acid molecules, expression systems, recombinant host cells, methods and kits, which enable the production and/or isolation and/or purification of polypeptides and untranslated RNA molecules. The compositions and methods of the invention can be applied in a regulatable expression system that is transiently transfected into mammalian host cells.

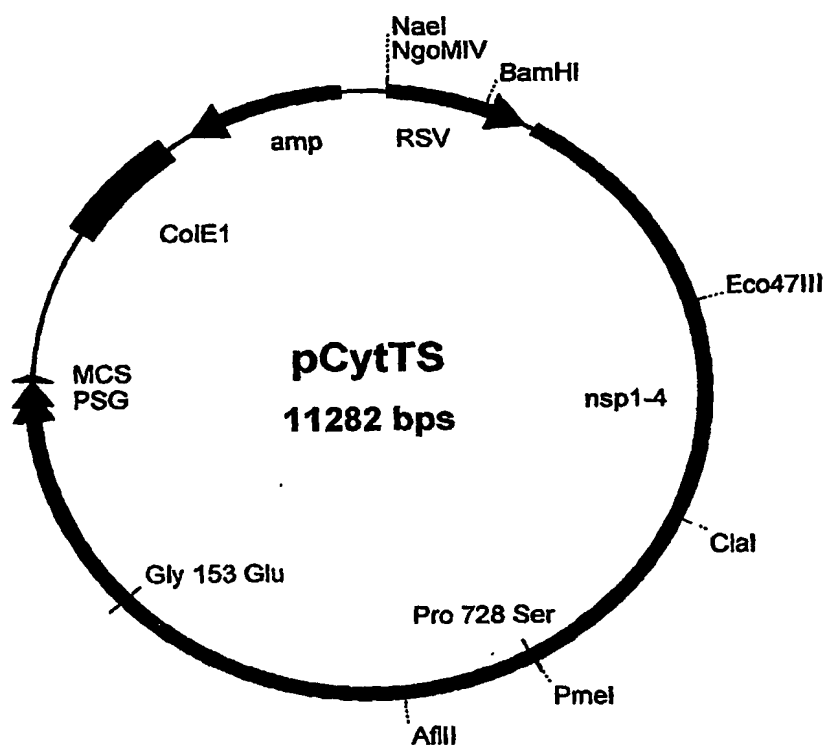


Fig. 1

2/9

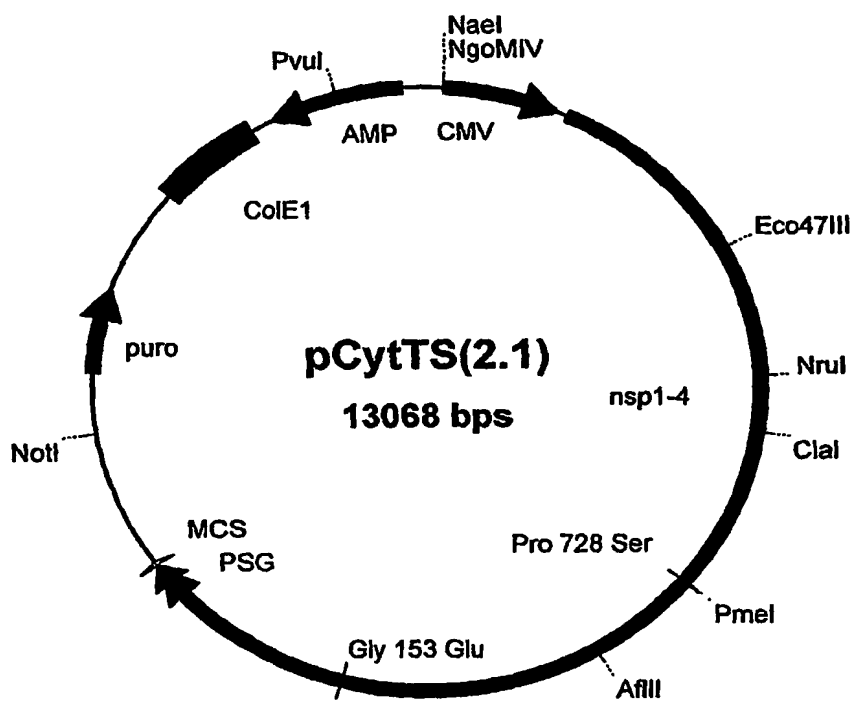
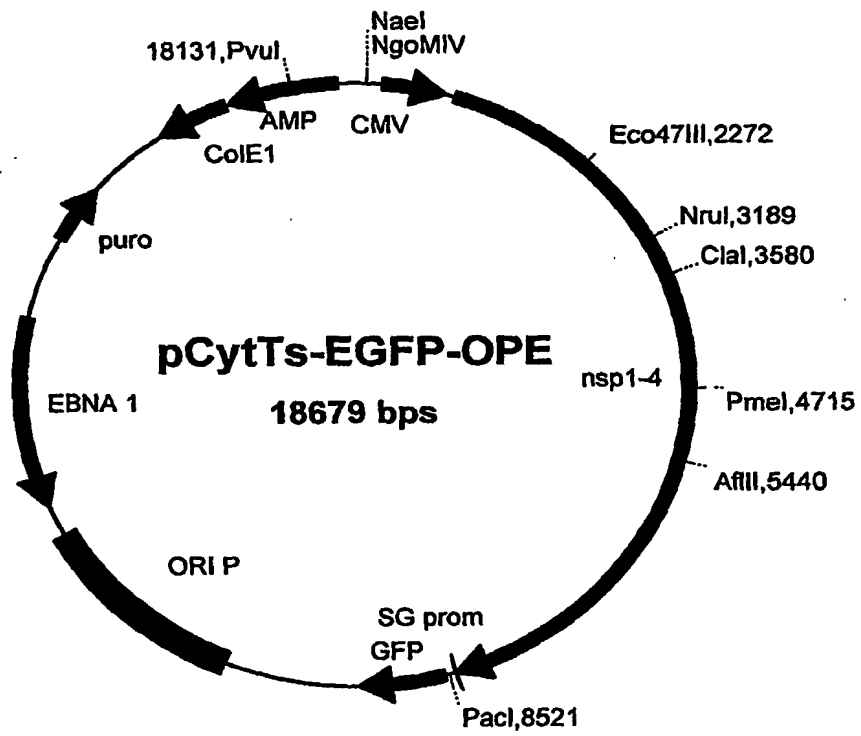


Fig. 2

3/9



**Fig. 3**



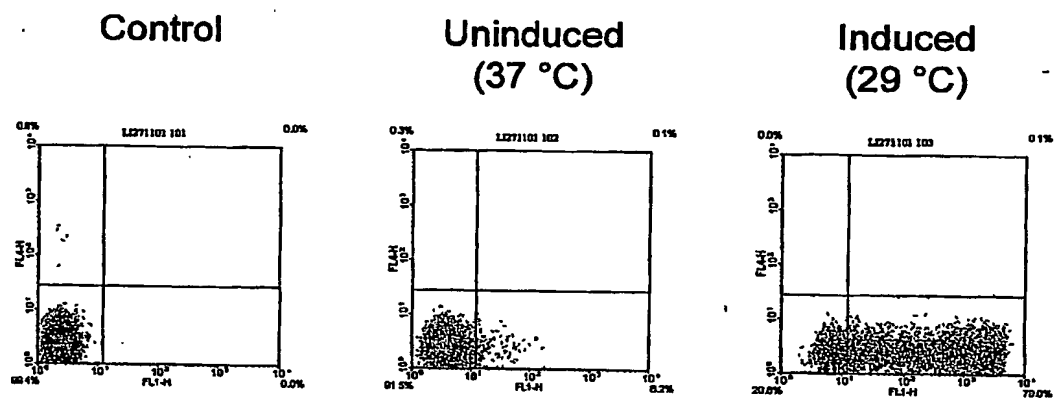


Fig. 4

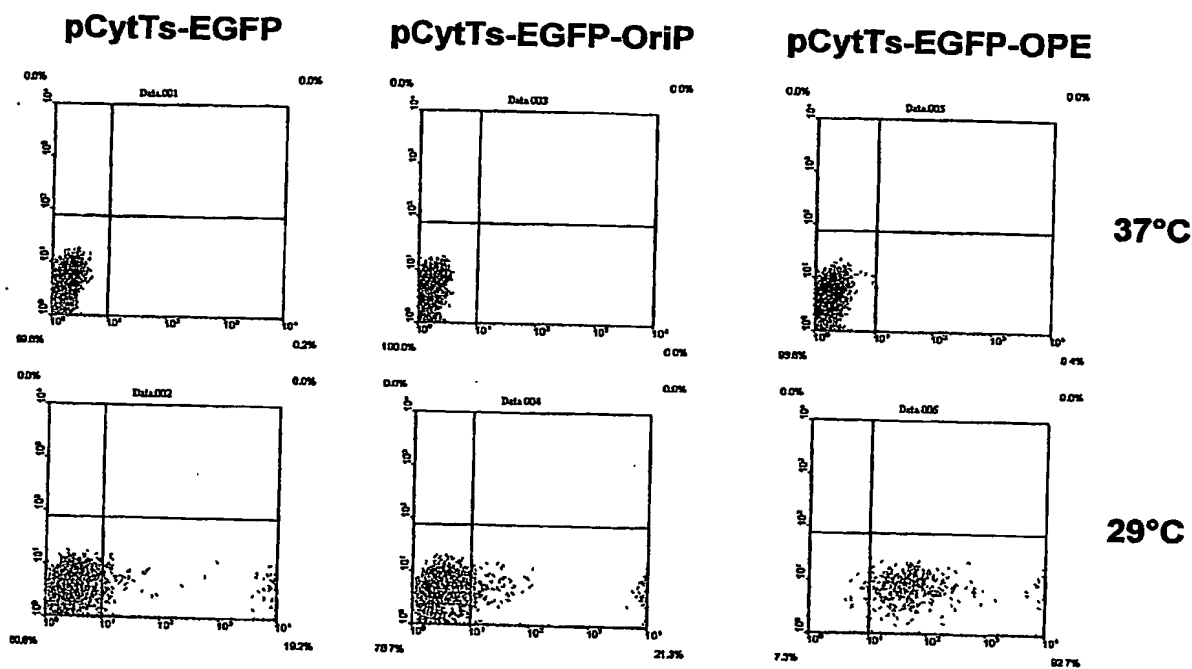
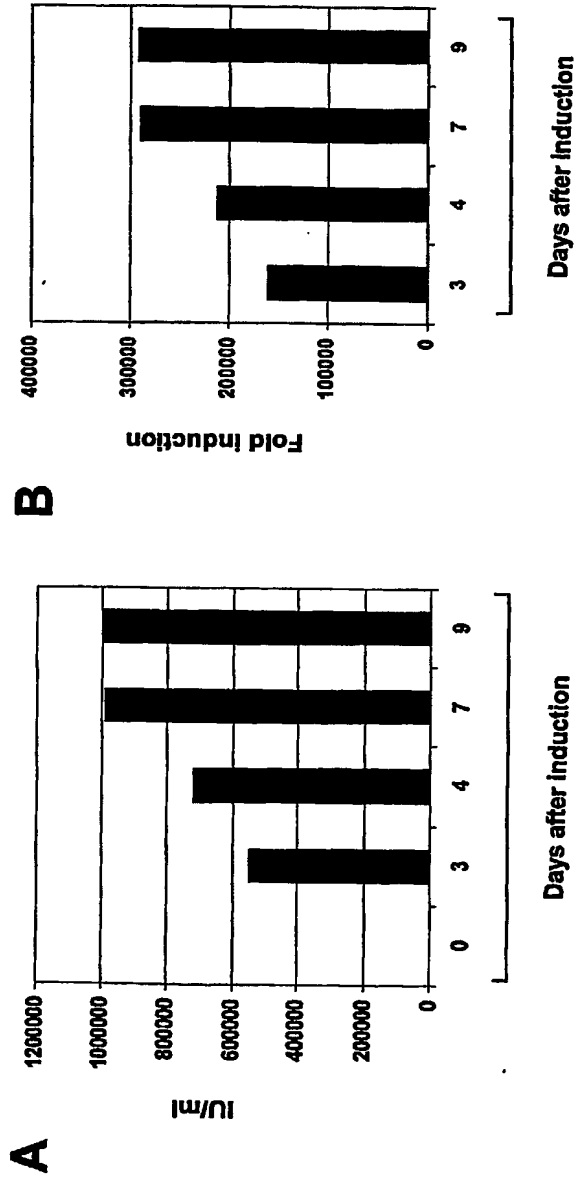


Fig. 5

6/9



**Fig. 6**

7/9

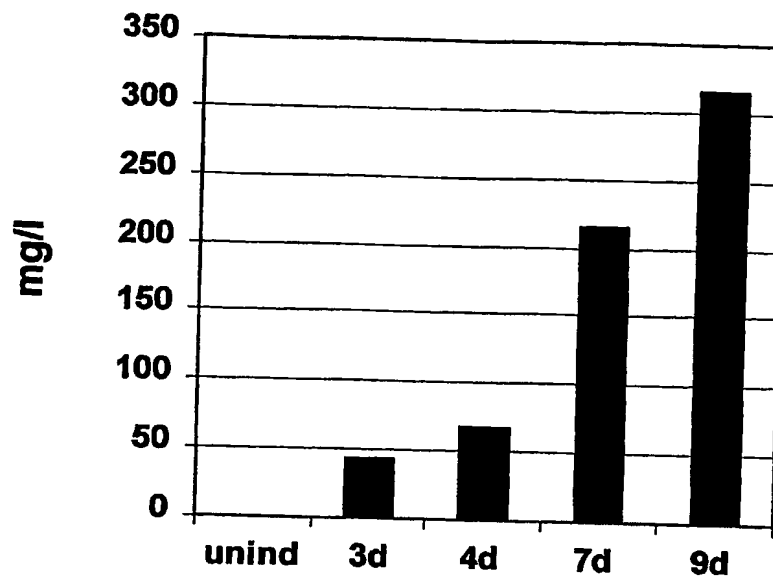
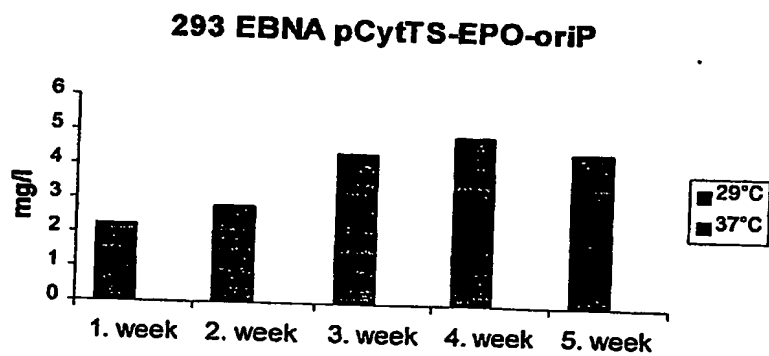


Fig. 7



9/9



**Fig. 9**

## SEQUENCE LISTING

**<110> Cytos Biotechnology AG**

**<120> Gene Expression System**

**<130> please provide**

**<160> 26**

**<170> PatentIn version 3.1**

**<210> 1**

**<211> 11282**

**<212> DNA**

**<213> Artificial Sequence**

**<220>**

**<223> CDNA**

**<400> 1**

ctgacgcgcc	ctgtagcggc	gcattaagcg	cggcggggtgt	gggtggttacg	cgcagcgtga	60
ccgctacact	tgccagcgcc	ctagcgcgcc	ctccttttcgc	tttcttccct	tcctttctcg	120
ccacgttcgc	cggctttccc	cgtcaagctc	taaatcgggg	gctcccttta	gggttccgat	180
ttagtgcttt	acggcacctc	gaccccaaaa	aacttgatta	gggtgatggg	tcacgtagtg	240
ggccatcgcc	ctgatagacg	gttttttcgcc	ctttgacggt	ggagtccacg	ttctttaata	300
gtggactctt	gttccaaact	ggaacaacac	tcaaccctat	ctcgggtctat	tcttttgatt	360
tataagggat	tttgccgatt	tcggcctatt	ggttaaaaaa	tgagctgatt	taacaaaaat	420
ttaacgcgaa	ttttaacaaa	atattaacgc	ttacaatttc	cattcgccat	tcaggctgcg	480
caactgttgg	gaagggcgat	cgggtgcgggc	ctcttcgcta	ttacgccagc	tggcgaaagg	540
gggatgtgct	gcaaggcgat	taagttgggt	aacgccaggg	ttttcccagt	cacgacgttg	600
taaaacgacg	gccagtgagc	gcgcaattaa	ccctcactaa	aggaacaaa	agctggctag	660
tggatccagt	cttatgcaat	actcttgtag	tcttgcaaca	tggtaacgat	gagttagcaa	720
catgccttac	aaggagagaa	aaagcaccgt	gcatgccgat	tggtggaagt	aagggtgtac	780
gatcgtgcct	tattaggaag	gcaacagacg	ggtctgacat	ggattggacg	aaccactgaa	840
ttccgcattg	cagagatatt	gtattttaagt	gccctacctc	gataccgtcg	agattgacgg	900
cgtagtacac	actattgaat	caaacagccg	accaattgca	ctaccatcac	aatggagaag	960
ccagtagtaa	acgtagacgt	agacccccag	agtccgtttg	tcgtgcaact	gcaaaaaaagc	1020
ttcccgcaat	ttgaggtagt	agcacagcag	gtcactccaa	atgaccatgc	taatgccaga	1080
gcatttttcgc	atctggccag	taaactaatc	gagctggagg	ttcctaccac	agcgacgatc	1140
ttggacatag	gcagcgcacc	ggctcgtaga	atgtttttccg	agcaccagta	tcatttgtgtc	1200

tgcccatgc gtagtccaga agacccggac cgcgatgatga aatacgccag taaactggcg	1260
gaaaaagcgt gcaagattac aaacaagaac ttgcatgaga agattaagga tctccggacc	1320
gtacttgata cgccggatgc tgaaacacca tcgctctgct ttcacaacga tgttacctgc	1380
aacatgcgtg ccgaatatc cgtcatgcag gacgtgtata tcaacgctcc cggaactatc	1440
tatcatcagg ctatgaaagg cgtgaggacc ctgtactgga ttggcttcga caccaccag	1500
ttcatgttct cggctatggc aggttcgtac cctgcgtaca acaccaactg ggccgacgag	1560
aaagtccttg aagcgcgtaa catcggactt tgcagcaca agctgagtga aggtaggaca	1620
ggaaaattgt cgataatgag gaagaaggag ttgaagccc ggtcgcgggt ttatttctcc	1680
gtaggatcga cactttatcc agaacacaga gccagcttgc agagctggca tcttccatcg	1740
gtgttccact tgaatggaaa gcagtcgtac acttgccgt gtgatacagt ggtgagttgc	1800
gaaggctacg tagtgaagaa aatcaccatc agtcccggga tcacgggaga aaccgtggga	1860
tacgcggtta cacacaatag cgagggcttc ttgctatgca aagttactga cacagtaaaa	1920
ggagaacggg tatcgttccc tgtgtgcacg tacatcccgg ccaccatatg cgatcagatg	1980
actggtataa tggccacgga tatatcacct gacgatgcac aaaaacttct ggttgggctc	2040
aaccagcgaa ttgtcattaa cggtaggact aacaggaaca ccaacaccat gcaaaattac	2100
cttctgccga tcatagcaca agggttcagc aaatgggcta aggagcgcaa ggatgatctt	2160
gataacgaga aaatgctggg tactagagaa cgcaagctta cgtatggctg cttgtggcg	2220
tttcgacta agaaagtaca ttcgttttat cgccacctg gaacgcagac ctgcgtaaaa	2280
gtcccgacct cttttagcgc ttttcccatg tcgtccgtat ggacgacctc tttgcccag	2340
tcgctgaggc agaaattgaa actggcattg caaccaaaga aggaggaaaa actgctgcag	2400
gtctcgagg aattagtcac ggaggccaag gctgcttttg aggatgctca ggaggaagcc	2460
agagcggaga agctccgaga agcacttcca ccattagtg cagacaaagg catcaggga	2520
gccgcagaag ttgtctgcga agtggagggg ctccaggcgg acatcggagc agcattagtt	2580
gaaacccgc gcggtcacgt aaggataata cctcaagcaa atgaccgtat gatcggacag	2640
tatatcgttg tctcgccaaa ctctgtgctg aagaatgcca aactcgcacc agcgcacccg	2700
ctagcagatc aggttaagat cataacacac tccggaagat caggaaggta cgcggtcgaa	2760
ccatacgacg cttaaagtact gatgccagca ggaggtgccg taccatggcc agaattccta	2820
gcactgagtg agagcgccac gttagtgtac aacgaaagag agtttgtgaa ccgcaaacta	2880
taccacattg ccatgcatgg ccccgccaag aatacagaag aggagcagta caaggttaca	2940
aaggcagagc ttgcagaaac agagtacgtg ttgacgtgg acaagaagcg ttgcgttaag	3000
aaggaagaag cctcaggtct ggtcctctcg ggagaactga ccaaccctcc ctatcatgag	3060
ctagctctgg agggactgaa gacccgacct gcggtcccgt acaaggtcga aacaatagga	3120
gtgataggca caccggggtc gggcaagtca gctattatca agtcaactgt caccgcacga	3180
gatcttgta ccagcggaaa gaaagaaaat tgtcgcgaaa ttgaggccga cgtgctaaga	3240



ctgaggggta	tgcagattac	gtcgaagaca	gtagattcgg	ttatgctcaa	cggatgccac	3300
aaagccgtag	aagtgctgta	cggtgacgaa	gcgttcgcgt	gccacgcagg	agcactactt	3360
gccttgattg	ctatcgtcag	gccccgcaag	aaggtagtac	tatgcggaga	cccatgcaa	3420
tgcggtattct	tcaacatgat	gcaactaaag	gtacatttica	atcacccctga	aaaagacata	3480
tgcaccaaga	cattctacaa	gtatatctcc	cggcggttgca	cacagccagt	tacagctatt	3540
gtatcgacac	tgcattacga	tggaaagatg	aaaaccacga	acccgtgcaa	gaagaacatt	3600
gaaatcgata	ttacaggggc	cacaaagccg	aagccagggg	atatcatcct	gacatgtttc	3660
cgcggttggg	ttaagcaatt	gcaaatcgac	tatcccggac	atgaagtaat	gacagccgcg	3720
gcctcacaag	ggctaaccag	aaaaggagtg	tatgccgtcc	ggcaaaaagt	caatgaaaac	3780
ccactgtacg	cgatcacatc	agagcatgtg	aacgtgttgc	tcacccgcac	tgaggacagg	3840
ctagtgtgga	aaaccttgca	gggcgaccca	tggattaagc	agcccactaa	catacctaaa	3900
ggaaactttc	aggctactat	agaggactgg	gaagctgaac	acaagggaat	aattgctgca	3960
ataaacagcc	ccactccccg	tgccaatccg	ttcagctgca	agaccaacgt	ttgctgggcg	4020
aaagcattgg	aaccgatact	agccacggcc	ggatcgttac	ttaccggttg	ccagtggagc	4080
gaactgttcc	cacagtttgc	ggatgacaaa	ccacattcgg	ccattttacgc	cttagacgta	4140
atttgcatta	agtttttcgg	catggacttg	acaagcggac	tgttttctaa	acagagcatc	4200
ccactaacgt	accatccccg	cgattcagcg	aggccggtag	ctcattggga	caacagccca	4260
ggaacccgca	agtatgggta	cgatcacgcc	attgccgccg	aactctcccc	tagattttccg	4320
gtgttccagc	tagctgggaa	gggcacacaa	cttgatttgc	agacggggag	aaccagagtt	4380
atctctgcac	agcataacct	ggtccccgtg	aaccgcaatc	ttcctcacgc	cttagtcccc	4440
gagtacaagg	agaagcaacc	cggccccgtc	aaaaaattct	tgaaccagtt	caaacaccac	4500
tcagtacttg	tggatatcaga	ggaaaaaatt	gaagctcccc	gtaagagaat	cgaatggatc	4560
gccccgattg	gcatagccgg	tgcatagata	aactacaacc	tggcttttcgg	gtttccgccg	4620
caggcacggt	acgacctggt	gttcatcaac	attggaacta	aatacagaaa	ccaccacttt	4680
cagcagtgcg	aagaccatgc	ggcgacctta	aaaacccttt	cgcgttcggc	cctgaattgt	4740
ttaaactcag	gaggcaccct	cgtggtgaag	tcctatggct	acgccgaccg	caacagtgcg	4800
gacgtagtca	ccgctcttgc	cagaaagttt	gtcagggtgt	ctgcagcgag	accagattgt	4860
gtctcaagca	atacagaaat	gtacctgatt	ttccgacaac	tagacaacag	ccgtacacgg	4920
caattcacc	cgcaccatct	gaattgcgtg	atttcgtccg	tgtatgaggg	tacaagagat	4980
ggagtggag	ccgcgccgtc	ataccgcacc	aaaaggggaga	atattgctga	ctgtcaagag	5040
gaagcagttg	tcaacgcagc	caatccgctg	ggtagaccag	gcgaaggagt	ctgccgtgcc	5100
atctataaac	gttggccgac	cagttttacc	gattcagcca	cggagacagg	caccgcaaga	5160
atgactgtgt	gcctaggaaa	gaaagtgatc	cacgcggtcg	gccctgattt	ccggaagcac	5220
ccagaagcag	aagccttgaa	attgctacaa	aacgcctacc	atgcagtggc	agacttagta	5280

aatgaacata acatcaagtc tgtcgccatt ccactgctat ctacaggcat ttacgcagcc 5340  
ggaaaagacc gccttgaagt atcacttaac tgcttgacaa ccgcgctaga cagaactgac 5400  
gcggacgtaa ccatctattg cctggataag aagtggaagg aaagaatcga cgcggcactc 5460  
caacttaagg agtctgtaac agagctgaag gatgaagata tggagatcga cgatgagtta 5520  
gtatggattc atccagacag ttgcttgaag ggaagaaagg gattcagtac tacaaaagga 5580  
aaattgtatt cgtacttcga aggcaccaa ttccatcaag cagcaaaaga catggcggag 5640  
ataaagggtcc tgttccctaa tgaccaggaa agtaatgaac aactgtgtgc ctacatattg 5700  
ggtgagacca tggaagcaat ccgcgaaaag tgcccggtcg accataaccc gtcgtctagc 5760  
ccgccccaaa cgttgccgtg cctttgcatg tatgccatga cgccagaaaag ggtccacaga 5820  
cttagaagca ataacgtcaa agaagttaca gtatgtcctt ccacccccct tcctaagcac 5880  
aaaattaaga atgttcagaa ggttcagtgc acgaaagtag tcctgtttaa tccgcacact 5940  
cccgcattcg ttcccggccc taagtacata gaagtgccag aacagcctac cgctcctcct 6000  
gcacaggccg aggaggcccc cgaagttgta gcgacaccgt caccatctac agctgataac 6060  
acctcgcttg atgtcacaga catctcactg gatatggatg acagtagcga aggctcactt 6120  
ttttcgagct ttagcggatc ggacaactct attactagta tggacagttg gtcgtcagga 6180  
cctagttcac tagagatagt agaccgaagg cagggtgggtg tggctgacgt tcatgccgtc 6240  
caagagcctg cccctattcc accgccaagg ctaaaagaaga tggcccgctt ggcagcggca 6300  
agaaaagagc ccactccacc ggcaagcaat agctctgagt ccctccacct ctcttttggt 6360  
ggggtatcca tgtccctcgg atcaattttc gacggagaga cggcccgcca ggcagcggta 6420  
caacccttg caacaggccc cacggatgtg cctatgtctt tcggatcgtt ttccgacgga 6480  
gagattgatg agctgagccg cagagtaact gagtccgaac ccgtcctggt tggatcattt 6540  
gaaccgggcg aagtgaactc aattatatcg tcccgatcag ccgtatcttt tccactacgc 6600  
aagcagagac gtagacgcag gagcaggagg actgaatact gactaaccgg ggtaggtggg 6660  
tacatatttt cgacggacac aggccctggg cacttgcaaa agaagtccgt tctgcagaac 6720  
cagcttacag aaccgacctt ggagcgcaat gtcctggaaa gaattcatgc cccggtgctc 6780  
gacacgtcga aagaggaaca actcaaactc aggtaccaga tgatgcccac cgaagccaac 6840  
aaaagtaggt accagtctcg taaagtagaa aatcagaaaag ccataaccac tgagcgacta 6900  
ctgtcaggac tacgactgta taactctgcc acagatcagc cagaatgcta taagatcacc 6960  
tatccgaaac cattgtactc cagtagcgta ccggcgaaact actccgatcc acagttcgct 7020  
gtagctgtct gtaacaacta tctgcatgag aactatccga cagtagcatc ttatcagatt 7080  
actgacgagt acgatgctta cttggatatg gtagacgaga cagtcgcatg cctggatact 7140  
gcaaccttct gccccgctaa gcttagaagt taccgaaaa aacatgagta tagagccccg 7200  
aatatccgca gtgcggttcc atcagcgatg cagaacacgc tacaaaatgt gtcattgcc 7260  
gcaactaaaa gaaattgcaa cgtcacgcag atgcgtgaac tgccaacact ggactcagcg 7320

acattcaatg	tcgaatgctt	tcgaaaatat	gcatgtaatg	acgagtat	ggagagagtc	7380
gctcggaagc	caattaggat	taccactgag	tttgtcaccg	catatgtagc	tagactgaaa	7440
ggccctaagg	ccgccgcact	atttgcaaag	acgtataatt	tgggtccatt	gcaagaagtg	7500
cctatggata	gattcgtcat	ggacatgaaa	agagacgtga	aagttacacc	aggcacgaaa	7560
cacacagaag	aaagaccgaa	agtacaagtg	atacaagccg	cagaaccctt	ggcgactgct	7620
tacttatgcg	ggattcaccg	ggaattagt	cgtaggctta	cggccgtctt	gcttccaaac	7680
attcacacgc	tttttgacat	gtcggcgagg	gattttgatg	caatcatagc	agaacacttc	7740
aagcaaggcg	acccgggtact	ggagacggat	atcgcatcat	tcgacaaaag	ccaagacgac	7800
gctatggcgt	taaccgggtct	gatgatcttg	gaggacctgg	gtgtggatca	accactactc	7860
gacttgatcg	agtgcgcctt	tggagaaata	tcattccacc	atctacctac	gggtactcgt	7920
tttaaattcg	gggcgatgat	gaaatccgga	atgttcctca	cactttttgt	caacacagtt	7980
ttgaatgtcg	ttatcgccag	cagagtacta	gaagagcggc	ttaaaacgtc	cagatgtgca	8040
gcgttcattg	gcgacgacaa	catcatacat	ggagtagtat	ctgacaaaga	aatggctgag	8100
aggtgcgcca	cctgggtcaa	catggagggt	aagatcatcg	acgcagtcac	cgggtgagaga	8160
ccaccttact	tctgcggcgg	atttatcttg	caagattcgg	ttacttccac	agcgtgccgc	8220
gtggcggaatc	ccctgaaaag	gctgtttaag	ttgggtaaac	cgctcccagc	cgacgacgag	8280
caagacgaag	acagaagacg	cgctctgcta	gatgaaacaa	aggcgtgggt	tagagtaggt	8340
ataacaggca	cttttagcagt	ggccgtgacg	acccgggtatg	aggtagacaa	tattcacctt	8400
gtcctactgg	cattgagaac	ttttgcccag	agcaaaagag	cattccaagc	catcagaggg	8460
gaaataaagc	atctctacgg	tggctcctaaa	tagtcagcat	agtagatctt	atctgactaa	8520
tactacaaca	ccaccacctc	tagacgcgta	gatctcacgt	gagcatgcag	gccttggggc	8580
caatgatccg	accagcaaaa	ctcgatgtac	ttccgaggaa	ctgatgtgca	taatgcatca	8640
ggctgggtaca	ttagatcccc	gcttaccgcg	ggcaatatag	caacactaaa	aactcgatgt	8700
acttccgagg	aagcgcagtg	cataatgctg	cgcagtggtg	ccacataacc	actatattaa	8760
ccatttatct	agcggacgcc	aaaaactcaa	tgtatttctg	aggaagcgtg	gtgcataatg	8820
ccacgcagcg	tctgcataac	ttttattatt	tcttttatta	atcaacaaaa	ttttgttttt	8880
aacatttcaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaggga	ttcccaactt	8940
gtttattgca	gcttataatg	gttacaaata	aagcaatagc	atcacaatt	tcacaaataa	9000
agcatttttt	tcactgcatt	ctagttgtgg	tttgtccaaa	ctcatcaatg	tatcttatca	9060
tgtctggatc	cgctgagacg	cgtccaattc	gccctatagt	gagtcgtatt	acgcgcgctt	9120
ggcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	9180
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaattgag	tgagctaact	9240
cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagct	9300
gcattaatga	atcggccaac	gcgcgggggag	aggcgggttg	cgtattgggc	gctcttccgc	9360

ac

<210> 2  
<211> 13068  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> cDNA

<400> 2  
ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga 60  
ccgctacact tgccagcgcc ctagcgcgcc ctcctttcgc tttcttcctt tcctttctcg 120  
ccacgttcgc cggcggatcc ggccattagc catattattc attgggtata tagcataaat 180  
caatattggc tattggccat tgcatacggt gtatccatat cataatatgt acattttatat 240  
tggctcatgt ccaacattac cgccatgttg acattgatta ttgactagtt attaataagta 300  
atcaattacg gggtcattag ttcataagccc atatattggag ttccgcgtta cataacttac 360  
ggtaaatggc ccgcctggct gaccgcccga cgacccccgc ccattgacgt caataatgac 420  
gtatgtttcc atagtaacgc caatagggac tttccattga cgtcaatggg tggagtattt 480  
acggtaaaact gccacttggc cagtacatca agtgtatcat atgccaagta cgccccctat 540  
tgacgtcaat gacggtaaat ggcccgcctg gcattatgcc cagtacatga ctttatggga 600  
ctttctact tggcagtaca tctacgtatt agtcatcgct attaccatgg tgatgcgggt 660  
ttggcagtac atcaatgggc gtggatagcg gtttgactca cggggatttc caagtctcca 720  
ccccattgac gtcaatggga gtttggtttg gcacaaaat caacgggact ttccaaaatg 780  
tcgtaacaac tccgccccat tgacgcaa at gggcggtagg cgtgtacggg gggagggtcta 840  
tataagcaga gtcgttttag tgaaccgtat tgacggcgta gtacacacta ttgaatcaaa 900  
cagccgacca attgcactac catcacaatg gagaagccag tagtaaacgt agacgtagac 960  
ccccagagtc cgtttgtcgt gcaactgcaa aaaagcttcc cgcaatttga ggtagtagca 1020  
cagcagggtca ctccaaatga ccatgcta at gccagagcat tttcgcatct ggccagtaaa 1080  
ctaatecagc tggagggttc taccacagcg acgatcttgg acataggcag cgcaccgggt 1140  
cgtagaatgt tttccgagca ccagtatcat tgtgtctgcc ccatgcgtag tccagaagac 1200  
ccggaccgca tgatgaaata tgccagtaaa ctggcggaaa aagcgtgcaa gattacaaac 1260  
aagaacttgc atgagaagat taaggatctc cggaccgtac ttgatacgcc ggatgctgaa 1320  
acaccatcgc tctgctttca caacgatgtt acctgcaaca tgcgtgccga atattccgtc 1380  
atgcaggacg tgtatatcaa cgctccccga actatctatc atcaggctat gaaaggcgtg 1440  
cggaccctgt actggattgg cttcgacacc acccagttca tgttctcggc tatggcaggt 1500  
tcgtaccctg cgtacaacac caactgggcc gacgagaaaag tccttgaagc gcgtaacatc 1560  
ggactttgca gcacaaagct gagtgaaggt aggacaggaa aattgtcgat aatgaggaag 1620  
aaggagtga agcccggtc gcgggtttat ttctccgtag gatcgacact ttatccagaa 1680  
cacagagcca gcttcagag ctggcatctt ccatcggtgt tccacttgaa tggaaagcag 1740  
tcgtacactt gccgctgtga tacagtgggt agttgcgaag gctacgtagt gaagaaaatc 1800

accatcagtc	ccgggatcac	gggagaaacc	gtgggatacg	cggttacaca	caatagcgag	1860
ggcttcttgc	tatgcaaagt	tactgacaca	gtaaaaggag	aacgggtatc	gttcctctgtg	1920
tgcacgtaca	tcccggccac	catatgcat	cagatgactg	gtataatggc	cacggatata	1980
tcacctgacg	atgcacaaaa	acttctgggt	gggctcaacc	agcgaattgt	cattaacggt	2040
aggactaaca	ggaacaccaa	caccatgcaa	aattaccttc	tgccgatcat	agcacaaggg	2100
ttcagcaaat	gggctaagga	gcgcaaggat	gatcttgata	acgagaaaaat	gctgggtact	2160
agagaacgca	agcttacgta	tggctgcttg	tgggcgtttc	gcactaagaa	agtacattcg	2220
ttttatcgcc	cacctggaac	gcagacctgc	gtaaaagtcc	cagcctcttt	tagcgctttt	2280
cccatgtcgt	ccgtatggac	gacctctttg	cccatgtcgc	tgaggcagaa	attgaaactg	2340
gcattgcaac	caaagaagga	ggaaaaactg	ctgcaggctc	cggaggaatt	agtcatggag	2400
gccaaaggctg	cttttgagga	tgctcaggag	gaagccagag	cggagaagct	ccgagaagca	2460
cttccaccat	tagtggcaga	caaaggcatc	gaggcagccg	cagaagttgt	ctgcgaagtg	2520
gaggggctcc	aggcggacat	cggagcagca	ttagttgaaa	ccccgcgcgg	tcacgtaagg	2580
ataatacctc	aagcaaatga	ccgtatgatc	ggacagtata	tcgttgcttc	gccaaactct	2640
gtgctgaaga	atgccaaact	cgaccacgcy	cacccgctag	cagatcaggt	taagatcata	2700
acacactccg	gaagatcagg	aagggtacgcy	gtcgaaccat	acgacgctaa	agtactgatg	2760
ccagcaggag	gtgccgtacc	atggccagaa	ttcctagcac	tgagtgagag	cgccacgtta	2820
gtgtacaacg	aaagagagtt	tgtgaaccgc	aaactatacc	acattgccat	gcatggcccc	2880
gccaaagaata	cagaagagga	gcagtacaag	gttacaagg	cagagcttgc	agaaacagag	2940
tacgtgtttg	acgtggacaa	gaagcgttgc	gttaagaagg	aagaagcctc	aggtctggtc	3000
ctctcgggag	aactgaccaa	ccctccctat	catgagctag	ctctggaggg	actgaagacc	3060
cgacctgcgg	tcccgtacaa	ggtcgaaaca	ataggagtga	taggcacacc	ggggtcgggc	3120
aagtcagcta	ttatcaagtc	aactgtcacg	gcacgagatc	ttgttaccag	cggaaagaaa	3180
gaaaattgtc	gcgaaattga	ggccgacgtg	ctaagactga	ggggtatgca	gattacgtcg	3240
aagacagtag	attcggttat	gctcaacgga	tgccacaaag	ccgtagaagt	gctgtacgtt	3300
gacgaagcgt	tcgcgtgcc	cgcaggagca	ctacttgctt	tgattgctat	cgtcaggccc	3360
cgcaagaagg	tagtactatg	cggagacccc	atgcaatgcg	gattcttcaa	catgatgcaa	3420
ctaaagggtac	atttcaatca	ccctgaaaaa	gacatatgca	ccaagacatt	ctacaagtat	3480
atctcccggc	gttgcacaca	gccagttaca	gctattgtat	cgacactgca	ttacgatgga	3540
aagatgaaaa	ccacgaaccc	gtgcaagaag	aacattgaaa	tcgatattac	agggggccaca	3600
aagccgaagc	caggggatat	catcctgaca	tgtttccgcg	ggtgggttaa	gcaattgcaa	3660
atcgactatc	ccggacatga	agtaatgaca	gcccgggcct	cacaagggct	aaccagaaaa	3720
ggagtgtatg	ccgtccggca	aaaagtcaat	gaaaaccac	tgtacgcgat	cacatcagag	3780
catgtgaacg	tgttgctcac	ccgcactgag	gacaggctag	tgtgaaaaac	cttgcaagggc	3840

gacccatgga	ttaagcagct	cactaacata	cctaaaggaa	actttcaggc	tactatagag	3900
gactgggaag	ctgaacacaa	gggaataatt	gctgcaataa	acagccccac	ttcccgtgcc	3960
aatccgttca	gctgcaagac	caacgtttgc	tgggcgaaag	catttgaacc	gatactagcc	4020
acggccggtg	tcgtacttac	cggttgccag	tggagcgaac	tgttcccaca	gtttgcggat	4080
gacaaaccac	attcggccat	ttacgcctta	gacgtaattt	gcattaagtt	tttcggcatg	4140
gacttgacaa	gcggactggt	ttctaaacag	agcatcccac	taacgtacca	ttccgccgat	4200
tcagcgaggc	cggtagctca	ttgggacaac	agcccaggaa	cccgcaagta	tgggtacgat	4260
cacgccattg	ccgccgaact	ctcccgtaga	tttccggtgt	tccagctagc	tgggaagggc	4320
acacaacttg	atttgcacac	ggggagaacc	agagtttatct	ctgcacagca	taacctggtc	4380
ccggtgaacc	gcaatcttcc	tcacgcctta	gtccccgagt	acaaggagaa	gcaaccggc	4440
ccggtcgaaa	aattcttgaa	ccagttcaaa	caccactcag	tacttgtggt	atcagaggaa	4500
aaaattgaag	ctccccgtaa	gagaatcgaa	tggatcgccc	cgattggcat	agccggtgca	4560
gataagaact	acaacctggc	tttcgggttt	ccgccgcagg	cacggtacga	cctggtgttc	4620
atcaacattg	gaactaaata	cagaaaccac	cactttcagc	agtgcgaaga	ccatgcggcg	4680
accttaaaaa	ccctttcgcg	ttcggccctg	aattgtttta	actcaggagg	caccctcgtg	4740
gtgaagtcct	atggctacgc	cgaccgcaac	agtgaggacg	tagtcaccgc	tcttgccaga	4800
aagtttgtca	gggtgtctgc	agcgagacca	gatttgtgtct	caagcaatac	agaaatgtac	4860
ctgattttcc	gacaactaga	caacagccgt	acacggcaat	tcaccccgca	ccatctgaat	4920
tgcgtgattt	cgcccggtga	tgagggtaca	agagatggag	ttggagccgc	gccgtcatac	4980
cgcaccaaaa	gggagaatat	tgctgactgt	caagaggaag	cagttgtcaa	cgcagccaat	5040
ccgctgggta	gaccaggcga	aggagtctgc	cgtgccatct	ataaacgttg	gccgaccagt	5100
tttaccgatt	cagccacgga	gacaggcacc	gcaägaatga	ctgtgtgcct	aggaaagaaa	5160
gtgatccacg	cggtcggccc	tgatttccgg	aagcaccacg	aagcagaagc	cttgaaattg	5220
ctacaaaacg	cctaccatgc	agtggcagac	ttagtaaattg	aacataacat	caagtctgtc	5280
gccattccac	tgctatctac	aggcatttac	gcagccggaa	aagaccgcct	tgaagtatca	5340
cttaactgct	tgacaaccgc	gctagacaga	actgacgcgg	acgtaaccat	ctattgcctg	5400
gataagaagt	ggaaggaaag	aatcgacgcg	gcactccaac	ttaaggagtc	tgtaacagag	5460
ctgaaggatg	aagatatgga	gatcgacgat	gagttagtat	ggattcatcc	agacagttgc	5520
ttgaagggaa	gaaagggtatt	cagtactaca	aaaggaaaat	tgtattcgta	cttcgaaggc	5580
accaaattcc	atcaagcagc	aaaagacatg	gcggagataa	aggtcctggt	ccctaattgac	5640
caggaaagta	atgaacaact	gtgtgcctac	atattgggtg	agaccatgga	agcaatccgc	5700
gaaaagtgcc	cggtcgacca	taacccgtcg	tctagcccg	ccaaaacggt	gccgtgcctt	5760
tgcatgtatg	ccatgacgcc	agaaagggtc	cacagactta	gaagcaataa	cgtcaaagaa	5820
gttacagtat	gctcctccac	cccccttcct	aagcacaana	ttaagaatgt	tcagaagggt	5880

cagtgcacga aagtagtcct gtttaatccg cacactcccg cattcgttcc cgcccgttaag	5940
tacatagaag tgccagaaca gcctaccgct cctcctgcac aggccgagga ggcccccgaa	6000
gttgtagcga caccgtcacc atctacagct gataacacct cgcttgatgt cacagacatc	6060
tcactggata tggatgacag tagcgaaggc tcactttttt cgagcttttag cggatcggac	6120
aactctatta ctagtatgga cagttggtcg tcaggaccta gttcactaga gatagtagac	6180
cgaaggcagg tgggtggtggc tgacgttcat gccgtccaag agcctgcccc tattccaccg	6240
ccaaggctaa agaagatggc ccgcctggca gcggcaagaa aagagcccac tccaccggca	6300
agcaatagct ctgagtcctt ccacctctct tttggtgggg tatccatgtc cctcggatca	6360
attttcgacg gagagacggc ccgccaggca gcggtacaac ccctggcaac agggcccacg	6420
gatgtgccta tgtctttcgg atcgttttcc gacggagaga ttgatgagct gagccgcaga	6480
gtaactgagt ccgaacccgt cctgttttga tcatttgaac cgggcgaagt gaactcaatt	6540
atatcgtccc gatcagccgt atcttttctt ctacgcaagc agagacgtag acgcaggagc	6600
aggaggactg aatactgact aaccggggta ggtgggtaca tattttcgac ggacacaggc	6660
cctgggcact tgcaaaagaa gtccgttctg cagaaccagc ttacagaacc gaccttgag	6720
cgcaatgtcc tggaaagaat tcatgccccg gtgctcgaca cgtcgaaaga ggaacaactc	6780
aaactcaggt accagatgat gccaccgaa gccaacaaaa gtaggtacca gtctcgtaaa	6840
gtagaaaatc agaaagccat aaccactgag cgactactgt caggactacg actgtataac	6900
tctgccacag atcagccaga atgctataag atcacctatc cgaaaccatt gtactccagt	6960
agcgtaccgg cgaactactc cgatccacag ttcgctgtag ctgtctgtaa caactatctg	7020
catgagaact atccgacagt agcatcttat cagattactg acgagtacga tgcttacttg	7080
gatatggtag acgagacagt cgcattgcctg gatactgcaa ctttctgccc cgctaagctt	7140
agaagttacc cgaaaaaaca tgagtataga gccccgaata tccgcagtgc ggttccatca	7200
gcgatgcaga acacgctaca aaatgtgtctc attgccgcaa ctaaaagaaa ttgcaacgtc	7260
acgcagatgc gtgaactgcc aacctggac tcagcgacat tcaatgtcga atgctttcga	7320
aaatatgcat gtaatgacga gtattgggag gagttcgctc ggaagccaat taggattacc	7380
actgagtttg tcaccgcata tgtagctaga ctgaaaggcc ctaaggccgc cgcactattt	7440
gcaaagacgt ataatttggt ccatttgcaa gaagtgccta tggatagatt cgtcatggac	7500
atgaaaagag acgtgaaagt tacaccaggc acgaaacaca cagaagaaag accgaaagta	7560
caagtgatac aagccgcaga acccctggcg actgcttact tatgcgggat tcaccgggaa	7620
ttagtgcgta ggcttacggc cgtcttgctt ccaaacattc acacgctttt tgacatgtcg	7680
gcggaggatt ttgatgcaat catagcagaa cacttcaagc aaggcgaccc ggtactggag	7740
acggatatcg catcattcga caaaagccaa gacgacgcta tggcgttaac cggctctgatg	7800
atcttgaggg acctgggtgt ggatcaacca ctactcgact tgatcgagtg cgcctttgga	7860
gaaatatcat ccaccatct acctacgggt actcgtttta aattcggggc gatgatgaaa	7920



tccggaatgt	tcctcacact	ttttgtcaac	acagttttga	atgtcgttat	cgccagcaga	7980
gtactagaag	agcggcttaa	aacgtccaga	tgtgcagcgt	tcattggcga	cgacaacatc	8040
atacatggag	tagtatctga	caaagaaatg	gctgagaggt	gcgccacctg	gctcaacatg	8100
gagggttaaga	tcatcgacgc	agtcacggt	gagagaccac	cttacttctg	cggcggattt	8160
atcttgcaag	attcggttac	ttccacagcg	tgccgcgtgg	cggatccccct	gaaaaggctg	8220
tttaagttgg	gtaaaccgct	cccagccgac	gacgagcaag	acgaagacag	aagacgcgct	8280
ctgctagatg	aaacaaaggc	gtggttttaga	gtagggtataa	caggcacttt	agcagtggcc	8340
gtgacgaccc	ggtatgaggt	agacaatatt	acacctgtcc	tactggcatt	gagaactttt	8400
gcccagagca	aaagagcatt	ccaagccatc	agaggggaaa	taaagcatct	ctacggtggt	8460
cctaaatagt	cagcatagta	catttcatct	gactaatacc	acaacaccac	cacctctaga	8520
ttaattaact	cgaggcgcgc	cgggcccaga	ggatctttgt	gaaggaacct	tacttctgtg	8580
gtgtgacata	attggacaaa	ctacctacag	agatttaaag	ctctaaggta	aatataaaat	8640
ttttaagtgt	ataatgtgtt	aaactactga	ttctaattgt	ttgtgtattt	tagattccaa	8700
cctatggaac	tgatgaatgg	gagcagtggg	ggaatgcctt	taatgaggaa	aacctgtttt	8760
gctcagaaga	aatgccatct	agtgatgatg	aggctactgc	tgactctcaa	cattctactc	8820
ctccaaaaaa	gaagagaaaag	gtagaagacc	ccaaggactt	tccttcagaa	ttgctaagtt	8880
ttttgagtca	tgctgtgttt	agtaatagaa	ctcttgcttg	ctttgctatt	tacaccacaa	8940
aggaaaaagc	tgactgtcta	tacaagaaaa	ttatggaaaa	atatttgatg	tagcggccca	9000
atgatccgac	cagcaaaaact	cgatgtactt	ccgaggaact	gatgtgcata	atgcatcagg	9060
ctggtacatt	agatccccgc	ttaccgcggg	caatatagca	acactaaaaa	ctcgatgtac	9120
ttccgaggaa	gcgcagtgcg	taatgctgcg	cagtgttgcc	acataaccac	tatattaacc	9180
atztatctag	cggacgccaa	aaactcaatg	tatttctgag	gaagcgtggg	gcataatgcc	9240
acgcagcgtc	tgcataactt	ttattatttc	ttttattaat	caacaaaatt	ttgtttttta	9300
catttcaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaag	ggaattccca	acttgtttat	9360
tgagcgttat	aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	9420
tttttactg	cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	atcatgtctg	9480
gatccgtcga	gacgcgtgcg	gccgcgttag	ggtgtggaaa	gtccccaggc	tccccagcag	9540
gcagaagtat	gcaaagcatg	catctcaatt	agtcagcaac	cagggtgtgga	aagtccccag	9600
gctccccagc	aggcagaagt	atgcaaagca	tgcatctcaa	ttagtcagca	accatagtcc	9660
cgcccctaac	tccgcccac	ccgcccctaa	ctccgcccag	ttccgcccac	tctccgcccc	9720
atggctgact	aatttttttt	atztatgcag	aggccgaggc	cgcttcggcc	tctgagctat	9780
tccagaagta	gtgaggaggc	ttttttggag	gcctaggctt	ttgcaaaaag	cttgcatgcc	9840
tgaggtcgg	ccgccacgac	cggtgccgcc	acatccccct	gaccacgcc	cctgaccctt	9900
cacaaggaga	cgaccttcca	tgaccgagta	caagcccacg	gtgcgcctcg	ccaccgcga	9960

cgacgtcccc cgggccgtac gcaccctcgc cgccgcgttc gccgactacc ccgccacgcg 10020  
ccacaccgtc gacccggacc gccacatcga gcgggtcacc gagctgcaag aactcttcct 10080  
cacgcgcgtc gggctcgaca tcggcaaggt gtgggtcgcg gacgacggcg ccgcggtggc 10140  
ggtctggacc acgccggaga gcgtcgaagc gggggcggtg ttcgccgaga tcggcccgcg 10200  
catggccgag ttgagcgggt cccggctggc cgcgagcaa cagatggaag gcctcctggc 10260  
gccgcaccgg cccaaggagc ccgcgtggtt cctggccacc gtcggcgtct cggccgacca 10320  
ccagggcaag ggtctgggca gcgccgtcgt gtcgcccgga gtggaggcgg ccgagcgcgc 10380  
cggggtgccc gccttcctgg agacctcgc gcccgcgaac ctccccttct acgagcggct 10440  
cggcttcacc gtcaccgccg acgtcgaggt gcccgaagga ccgcgcacct ggtgcatgac 10500  
ccgcaagccc ggtgcctgac gccgcgcccga cgaccgcag cggccgaccg aaaggagcgc 10560  
acgaccccat ggctccgacc gaagccaccc ggggcggccc cgccgacccc gcacccgccc 10620  
ccgaggccca ccgactctag ctagaggatc ataactcagc ataccacatt ttagagaggt 10680  
ttacttgctt taaaaaacct ccacacctc cccctgaacc tgaaacataa aatgaatgca 10740  
attgttggtt ttaacttggt tattgcagct tataatgggt acaaataaag caatagcatc 10800  
acaaatttca caaataaagc atttttttca ctgcattcta gttgtgggtt gtccaacgcg 10860  
tcggtaccag cttttgttcc ctttagtgag ggttaatttc gagcttggcg taatcatggt 10920  
catagctggt tcctgtgtga aattgttatc cgctcacaat tccacacaac atacgagccg 10980  
gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca ttaattgcgt 11040  
ttcgctcact gcccgcttcc cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg 11100  
gccaacgcgc ggggagaggc gggttgcgta ttgggcgctc ttccgcttcc tcgctcactg 11160  
actcgtcgcg ctcggtcgtt cggctgcggc gagcggtatc agctcactca aaggcggtaa 11220  
tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc 11280  
aaaaggccag gaaccgtaaa aaggccgcgt tgctggcggt tttccatagg ctccgcccc 11340  
ctgacgagca tcacaaaaat cgacgtcaa gtcagaggtg gcgaaacccg acaggactat 11400  
aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctggt ccgaccctgc 11460  
cgcttaccgg atacctgtcc gcctttctcc cttcggaag cgtggcgctt tctcatagct 11520  
cacgtgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg 11580  
aacccccgt tcagcccgcg cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc 11640  
cggtgaagaca cgacttatcg ccaactggcag cagccactgg taacaggatt agcagagcga 11700  
ggtatgtagg cgggtgctaca gagttcttga agtgggtggc taactacggc tacactagaa 11760  
gaacagtatt tgggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta 11820  
gctcttgatc cggcaaacaa accaccgctg gtagcgggtg tttttttggt tgcaagcagc 11880  
agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct acgggggtctg 11940  
acgctcagtg gaacgaaaac tcacgttaag ggattttggt catgagatta tcaaaaagga 12000

tcttcaccta gatcctttta aattaataaat gaagtttttaa atcaatctaa agtatatatg 12060  
 agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc tcacgcatct 12120  
 gtctattttcg ttcattccata gttgcctgac tccccgctcg gtagataact acgatacggg 12180  
 agggccttacc atctggcccc agtgctgcaa tgataccgcg agaccacgcg tcaccggctc 12240  
 cagatttatc agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa 12300  
 ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc 12360  
 cagttaatag tttgcgcaac gttgttgcca ttgctacagg catcgtggtg tcacgctcgt 12420  
 cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt acatgatccc 12480  
 ccatgttggtg caaaaaagcg gttagctcct tcggctcctc gatcgttggtc agaagtaagt 12540  
 tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc 12600  
 catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt 12660  
 gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg ggataatacc gcgccacata 12720  
 gcagaacttt aaaagtgtc atcattggaa aacgtttctt ggggcgaaaa ctctcaagga 12780  
 tcttaccgct gttgagatcc agttcgatgt aaccactcgc tgcacccaac tgatcttcag 12840  
 catctttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa atgcccga 12900  
 aaaagggaat aaggcgaca cggaaatgtt gaatactcat actcttcctt tttcaatatt 12960  
 attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga 13020  
 aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccac 13068

<210> 3  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> primer

<400> 3  
 aacattgaaa tcgatattac agggg

25

<210> 4  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> primer

<400> 4  
 cgggttatgg tcgaccgggc

20

<210> 5  
 <211> 40  
 <212> DNA  
 <213> Artificial Sequence

14

<220>  
<223> primer

<400> 5  
gtgccctccc ctgagtttaa acaattcagg gccgaacgcg

40

<210> 6  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 6  
gaattgttta aactcaggag gcaccctcgt gg

32

<210> 7  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 7  
ggtagacgag acagtcgcat gcctggatac

30

<210> 8  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 8  
gtatccaggc atgcgactgt ctcgtctacc

30

<210> 9  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 9  
cagaccggtt aacgccatag cgtcg

25

<210> 10  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 10  
ctctattact agtatggaca gttgg

25

<210>	11
<211>	27
<212>	DNA
<213>	Artificial sequence

**<220>**  
**<223> primer**

**<400> 11**  
**ctagattaat taactcgagg cgcgccg**

<210>	12
<211>	27
<212>	DNA
<213>	Artificial sequence

**<220>**  
**<223> primer**

**<400> 12**  
**ggcccggcgc gcctcgagtt aattaat**

<210>	13
<211>	30
<212>	DNA
<213>	Artificial Sequence

**<220>**  
**<223>** primer

**<400> 13**  
**ataagaatgc cggcggatcc ggccattagc**

<b>&lt;210&gt;</b>	<b>14</b>
<b>&lt;211&gt;</b>	<b>42</b>
<b>&lt;212&gt;</b>	<b>DNA</b>
<b>&lt;213&gt;</b>	<b>Artificial sequence</b>

**<220>**  
**<223> primer**

<400> 14  
ccgtcaatac ggttcactaa acgagctctg cttatataga cc

<b>&lt;210&gt;</b>	<b>15</b>
<b>&lt;211&gt;</b>	<b>38</b>
<b>&lt;212&gt;</b>	<b>DNA</b>
<b>&lt;213&gt;</b>	<b>Artificial sequence</b>

**<220>**  
**<223>** primer

**<400> 15**  
gctcgttttag tgaaccgtat tgacggcgta gtacacac

<b>&lt;210&gt;</b>	<b>16</b>
<b>&lt;211&gt;</b>	<b>20</b>
<b>&lt;212&gt;</b>	<b>DNA</b>
<b>&lt;213&gt;</b>	<b>Artificial Sequence</b>

**<220>**  
**<223> primer**

<400> 16  
acgtcggcct caatttcgcg 20

<210> 17  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 17  
gcgcgcgggc ccagaggatc tttgtgaagg 30

<210> 18  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 18  
gcgcgcgcgg ccgctacatc aaatattttt cc 32

<210> 19  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 19  
acgtacgcgt gcggccgcgt taggggtgtgg aaagtcccc 39

<210> 20  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 20  
acgtacgcgt tggacaaacc acaactagaa tgc 33

<210> 21  
<211> 15081  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> cDNA

<400> 21  
ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga 60  
ccgctacact tgccagcgcc ctacgcgccg ctcttttcgc tttcttccct tcctttctcg 120  
ccacgttcgc cggcggatcc ggccattagc catattattc attgggtata tagcataaat 180  
caatattggc tattggccat tgcatacgtt gtatccatat cataatatgt acatttatat 240

tggtcatgt ccaacattac cgccatgttg acattgatta ttgactagtt attaatagta	300
atcaattacg gggtcattag ttcataagccc atatatggag ttccgcgtta cataacttac	360
ggtaaattggc ccgcctggct gaccgccccaa cgacccccgc ccattgacgt caataatgac	420
gtatgttccc atagtaacgc caatagggac tttccattga cgtcaatggg tggagtattt	480
acggtaaact gcccacttgg cagtacatca agtgtatcat atgccaagta cgccccctat	540
tgacgtcaat gacggtaaat ggcccgctg gcattatgcc cagtacatga ctttatggga	600
ctttcctact tggcagtaca tctacgtatt agtcatcgct attaccatgg tgatgcgggt	660
ttggcagtac atcaatgggc gtggatagcg gtttgactca cggggatttc caagtctcca	720
ccccattgac gtcaatggga gtttggtttg gcacaaaaat caacgggact ttccaaaatg	780
tcgtaacaac tccgccccat tgacgcaaat gggcggtagg cgtgtacggg gggaggtcta	840
tataagcaga gctcgtttag tgaaccgtat tgacggcgta gtacacacta ttgaatcaaa	900
cagccgacca attgcactac catcacaatg gagaagccag tagtaaacgt agacgtagac	960
ccccagagtc cgtttgtcgt gcaactgcaa aaaagcttcc cgcaatttga ggtagtagca	1020
cagcagggtca ctccaaatga ccattgcta ggcagagcat ttctcgatct ggccagtaaa	1080
ctaactcgagc tggaggttcc taccacagcg acgatcttgg acataggcag cgcaccggct	1140
cgtagaatgt tttccgagca ccagtatcat tgtgtctgcc ccatgcgtag tccagaagac	1200
ccggaccgca tgatgaaata tgccagtaaa ctggcgga aaagcgtgcaa gattacaaac	1260
aagaacttgc atgagaagat taaggatctc cggaccgtac ttgatacgcc ggatgctgaa	1320
acaccatcgc tctgctttca caacgatgtt acctgcaaca tgcgtgccga atattccgtc	1380
atgcaggacg tgtatatcaa cgctcccga actatctatc atcaggctat gaaaggcgtg	1440
cggaccctgt actggattgg cttcgacacc acccagttca tgttctcggc tatggcaggt	1500
tcgtaccctg cgtacaacac caactgggcc gacgagaaag tccttgaagc gcgtaacatc	1560
ggactttgca gcacaaagct gagtgaagg aggacaggaa aattgtcgat aatgaggaag	1620
aaggagtga agcccggtc gcgggtttat ttctccgtag gatcgacact ttatccagaa	1680
cacagagcca gcttgcagag ctggcatctt ccatcggtgt tccacttgaa tggaaagcag	1740
tcgtacactt gccgctgtga tacagtgggt agttgcgaag gctacgtagt gaagaaaatc	1800
accatcagtc ccgggatcac gggagaaacc gtgggatacg cggttacaca caatagcgag	1860
ggcttcttgc tatgcaaagt tactgacaca gtaaaaggag aacgggtatc gttccctgtg	1920
tgacgtaca tcccggccac catatgcgat cagatgactg gtataatggc cacggatata	1980
tcacctgacg atgcacaaaa acttctgggt gggctcaacc agcgaattgt cattaacggt	2040
aggactaaca ggaacaccaa caccatgcaa aattaccttc tgccgatcat agcacaaggg	2100
ttcagcaaat gggctaagga gcgcaaggat gatcttgata acgagaaaat gctgggtact	2160
agagaacgca agcttacgta tggctgcttg tgggcgtttc gactaagaa agtacattcg	2220
ttttatcgcc cacctggaac gcagacctgc gtaaaagtcc cagcctcttt tagcgctttt	2280

cccatgtcgt	ccgtatggac	gacctctttg	cccatgtcgc	tgaggcagaa	attgaaactg	2340
gcattgcaac	caaagaagga	ggaaaaactg	ctgcaggctt	cggaggaatt	agtcattggag	2400
gccaaggctg	cttttgagga	tgctcaggag	gaagccagag	cggagaagct	ccgagaagca	2460
cttccaccat	tagtggcaga	caaaggcatc	gaggcagccg	cagaagttgt	ctgcgaagtg	2520
gaggggctcc	aggcggacat	cggagcagca	ttagttgaaa	ccccgcgcgg	tcacgtaagg	2580
ataatacctc	aagcaaatga	ccgtatgatc	ggacagtata	tcgttgcttc	gccaaactct	2640
gtgctgaaga	atgccaaact	cgcaccagcg	cacccgctag	cagatcaggt	taagatcata	2700
acacactccg	gaagatcagg	aagggtacgc	gtcgaaccat	acgacgctaa	agtactgatg	2760
ccagcaggag	gtgccgtacc	atggccagaa	ttcctagcac	tgagtgagag	cgccacgtta	2820
gtgtacaacg	aaagagagtt	tgtgaaccgc	aaactatacc	acattgccat	gcatggcccc	2880
gccaagaata	cagaagagga	gcagtacaag	gttacaagg	cagagcttgc	agaaacagag	2940
tacgtgtttg	acgtggacaa	gaagcgttgc	gttaagaagg	aagaagcctc	aggtctggtc	3000
ctctcgggag	aactgaccaa	ccctccctat	catgagctag	ctctggaggg	actgaagacc	3060
cgacctgcgg	tcccgtacaa	ggtcgaaaca	ataggagtga	taggcacacc	ggggtcgggc	3120
aagtcagcta	ttatcaagtc	aactgtcacg	gcacgagatc	ttgttaccag	cggaaagaaa	3180
gaaaattgtc	gcgaaattga	ggccgacgtg	ctaagactga	ggggatatga	gattacgtcg	3240
aagacagtag	attcggttat	gtcaacgga	tgccacaaag	ccgtagaagt	gctgtacgtt	3300
gacgaagcgt	tcgctgcca	cgcaggagca	ctacttgctt	tgattgctat	cgtcaggccc	3360
cgcaagaagg	tagtactatg	cggagacccc	atgcaatgcg	gattcttcaa	catgatgcaa	3420
ctaaaggtag	atttcaatca	ccctgaaaaa	gacatatgca	ccaagacatt	ctacaagtat	3480
atctcccggc	gttgacaca	gccagttaca	gctattgtat	cgacactgca	ttacgatgga	3540
aagatgaaaa	ccacgaaccc	gtgcaagaag	aacattgaaa	tcgatattac	aggggccaca	3600
aagccgaagc	caggggatat	catcctgaca	tgtttccgcg	gggtgggttaa	gcaattgcaa	3660
atcgactatc	ccggacatga	agtaatgaca	gccgcggcct	cacaagggct	aaccagaaaa	3720
ggagtgtatg	ccgtccggca	aaaagtcaat	gaaaaccac	tgtacgcgat	cacatcagag	3780
catgtgaacg	tggtgctcac	ccgactgag	gacaggctag	tggtggaaaac	cttgaggggc	3840
gacccatgga	ttaagcagct	cactaacata	cctaaaggaa	actttcaggc	tactatagag	3900
gactgggaag	ctgaacacaa	gggaataatt	gctgcaataa	acagccccac	tccccgtgcc	3960
aatccgttca	gctgcaagac	caacgtttgc	tgggcgaaag	cattggaacc	gatactagcc	4020
acggccggta	tcgtacttac	cggttgccag	tggagcgaac	tgttcccaca	gtttgcggat	4080
gacaaaccac	attcggccat	ttacgcctta	gacgtaattt	gcattaagtt	tttcggcatg	4140
gacttgacaa	gcggactgtt	ttctaaacag	agcatccac	taacgtacca	tcccgcgat	4200
tcagcgaggc	cggtagctca	ttgggacaac	agcccaggaa	cccgaagta	tgggtacgat	4260
cacgccattg	ccgccgaact	ctcccgtaga	tttccggtgt	tccagctagc	tgggaagggc	4320



acacaacttg	atttgcagac	ggggagaacc	agagttatct	ctgcacagca	taacctggtc	4380
ccggtgaacc	gcaatcttcc	tcacgcctta	gtccccgagt	acaaggagaa	gcaacccggc	4440
ccggtcga	aaattcttga	ccagttcaaa	caccactcag	tacttgtggt	atcagaggaa	4500
aaaattgaag	ctccccgtaa	gagaatcgaa	tggatcgccc	cgattggcat	agccggtgca	4560
gataagaact	acaacctggc	tttcggggtt	ccgccgcagg	cacggtacga	cctgggtgtc	4620
atcaacattg	gaactaaata	cagaaaccac	cactttcagc	agtgcgaaga	ccatgcggcg	4680
accttaaaaa	ccctttcgcg	ttcggccctg	aattgtttaa	actcaggagg	caccctcgtg	4740
gtgaagtcct	atggctacgc	cgaccgcaac	agtgaggacg	tagtcaccgc	tcttgccaga	4800
aagtttgtca	gggtgtctgc	agcgagacca	gattgtgtct	caagcaatac	agaaatgtac	4860
ctgattttcc	gacaactaga	caacagccgt	acacggcaat	tcaccccgca	ccatctgaat	4920
tgcgtgattt	cgtccgtgta	tgagggtaca	agagatggag	ttggagccgc	gccgtcatac	4980
cgcacaaaaa	gggagaatat	tgctgactgt	caagaggaag	cagttgtcaa	cgcagccaat	5040
ccgctgggta	gaccaggcga	aggagtctgc	cgtgccatct	ataaacgttg	gccgaccagt	5100
tttaccgatt	cagccacgga	gacaggcacc	gcaagaatga	ctgtgtgcct	aggaaagaaa	5160
gtgatccacg	cggctggccc	tgattttcgg	aagcaccag	aagcagaagc	cttgaaattg	5220
ctacaaaacg	cctaccatgc	agtggcagac	ttagtaaatg	aacataacat	caagtctgtc	5280
gccattccac	tgctatctac	aggcatttac	gcagccggaa	aagaccgcct	tgaagtatca	5340
cttaactgct	tgacaaccgc	gctagacaga	actgacgcgg	acgtaaccat	ctattgcctg	5400
gataagaagt	ggaaggaaag	aatcgacgcg	gcactccaac	ttaaggagtc	tgtaacagag	5460
ctgaaggatg	aagatatgga	gatcgacgat	gagttagtat	ggattcatcc	agacagttgc	5520
ttgaagggaa	gaaagggatt	cagtactaca	aaaggaaaat	tgtattcgta	cttcgaaggc	5580
accaaattcc	atcaagcagc	aaaagacatg	gcggagataa	aggtcctgtt	ccctaattgac	5640
caggaaagta	atgaacaact	gtgtgcctac	atattgggtg	agaccatgga	agcaatccgc	5700
gaaaagtgcc	cggctgacca	taaccctgcg	tctagcccgc	ccaaaacgtt	gccgtgcctt	5760
tgcatgtatg	ccatgacgcc	agaaaggggtc	cacagactta	gaagcaataa	cgtcaaagaa	5820
gttacagtat	gctcctccac	cccccttcct	aagcacaaaa	ttaagaatgt	tcagaagggtt	5880
cagtgcacga	aagtagtcct	gtttaatccg	cacactcccc	cattcgttcc	cgcccgtaa	5940
tacatagaag	tgccagaaca	gcctaccgct	cctcctgcac	aggccgagga	ggcccccgaa	6000
gttgtagcga	caccgtcacc	atctacagct	gataaacact	cgcttgatgt	cacagacatc	6060
tcactggata	tggatgacag	tagcgaaggc	tcactttttt	cgagcttttag	cggatcggac	6120
aactctatta	ctagtatgga	cagttgggtc	tcaggacctt	gttcactaga	gatagtagac	6180
cgaaggcagg	tggtgggtgg	tgacgttcat	gccgtccaag	agcctgcccc	tattccaccg	6240
ccaaggctaa	agaagatggc	ccgcctggca	gcggcaagaa	aagagcccac	tccaccggca	6300
agcaatagct	ctgagtcctt	ccacctctct	tttgggtggg	tatccatgtc	cctcggatca	6360

attttcgacg gagagacggc ccgccaggca gcggtacaac ccctggcaac agggcccccacg	6420
gatgtgccta tgtcttttcgg atcgttttcc gacggagaga ttgatgagct gagccgcaga	6480
gtaactgagt ccgaaccctgt cctgttttga tcatttgaac cgggcgaagt gaactcaatt	6540
atatcgcccc gatcagccgt atctttttcct ctacgcaagc agagacgtag acgcaggagc	6600
aggaggactg aatactgact aaccggggta ggtgggtaca tatttttcgac ggacacaggc	6660
cctgggcaact tgcaaaagaa gtccgtttctg cagaaccagc ttacagaacc gaccttggag	6720
cgcaatgtcc tggaaagaat tcatgccccg gtgctcgaca cgtcgaaaga ggaacaactc	6780
aaactcaggt accagatgat gccaccgaa gccaaacaaa gtaggtacca gtctcgtaaa	6840
gtagaaaatc agaaagccat aaccactgag cgactactgt caggactacg actgtataac	6900
tctgccacag atcagccaga atgctataag atcacctatc cgaaaccatt gtactccagt	6960
agcgtaccgg cgaactactc cgatccacag ttgctgttag ctgtctgtaa caactatctg	7020
catgagaact atccgacagt agcatcttat cagattactg acgagtacga tgcttacttg	7080
gatatggtag acgagacagt cgcattgcctg gatactgcaa ctttctgccc cgctaagctt	7140
agaagttacc cgaaaaaaca tgagtataga gccccgaata tccgcagtgc ggttccatca	7200
gcgatgcaga acacgctaca aaatgtgctc attgccgcaa ctaaaagaaa ttgcaacgctc	7260
acgcagatgc gtgaactgcc aacactggac tcagcgacat tcaatgtcga atgctttcga	7320
aaatatgcat gtaatgacga gtattgggag gagttcgctc ggaagccaat taggattacc	7380
actgagtttg tcaccgcata tgtagctaga ctgaaaggcc ctaaggccgc cgcactattt	7440
gcaaagacgt ataatttggt cccattgcaa gaagtgccta tggatagatt cgtcatggac	7500
atgaaaagag acgtgaaagt tacaccaggc acgaaacaca cagaagaaag accgaaagta	7560
caagtgatac aagccgcaga acccctggcg actgcttact tatgcgggat tcaccgggaa	7620
ttagtgcgta ggcttacggc cgtcttgctt ccaaacattc acacgctttt tgacatgtcg	7680
gcggaggatt ttgatgcaat catagcagaa cacttcaagc aaggcgaccc ggtactggag	7740
acggatatcg catcattcga caaaagccaa gacgacgcta tggcgtaaac cggctctgatg	7800
atcttggagg acctgggtgt ggatcaacca ctactcgact tgatcgagtg cgccttttga	7860
gaaatatcat ccacccatct acctacgggt actcgtttta aattcggggc gatgatgaaa	7920
tccggaatgt tcctcacact ttttgtcaac acagttttga atgtcgttat cgccagcaga	7980
gtactagaag agcggcttaa aacgtccaga tgtgcagcgt tcattggcga cgacaacatc	8040
atacatggag tagtatctga caaagaaatg gctgagaggt gcgccacctg gctcaacatg	8100
gaggtttaaga tcatcgacgc agtcatcggt gagagaccac cttacttctg cggcggatgt	8160
atcttgcaag attcggttac ttccacagcg tgccgcgtgg cggatccctt gaaaaggctg	8220
tttaagttgg gtaaaccgct cccagccgac gacgagcaag acgaagacag aagacgcgct	8280
ctgctagatg aaacaaaggc gtggttttaga gtaggtataa caggcacttt agcagtggcc	8340
gtgacgaccc ggtatgaggt agacaatatt acacctgtcc tactggcatt gagaactttt	8400

gcccagagca	aaagagcatt	ccaagccatc	agaggggaaa	taaagcatct	ctacggtggt	8460
cctaaatagt	cagcatagta	catttcatct	gactaatacc	acaacaccac	ca'ctctaga	8520
ttaattaact	cgaggcgcgc	cgggcccgaga	ggatctttgt	gaaggaacct	tacttctgtg	8580
gtgtgacata	attggacaaa	ctacctacag	agatttaaag	ctctaaggta	aatataaaat	8640
ttttaagtgt	ataatgtgtt	aaactactga	ttctaattgt	ttgtgtat'tt	tagattccaa	8700
cctatggaac	tgatgaatgg	gagcagtggt	ggaatgcctt	taatgaggaa	aacctgtttt	8760
gctcagaaga	aatgccatct	agtgatgatg	aggctactgc	tgactctcaa	cattctactc	8820
ctccaaaaaa	gaagagaaaag	gtagaagacc	ccaaggactt	tccttcagaa	ttgctaagtt	8880
ttttgagtca	tgctgtgttt	agtaatagaa	ctcttgcttg	ctttgctatt	tacaccacaa	8940
aggaaaaagc	tgactgcta	tacaagaaaa	ttatggaaaa	atatttgatg	tagcggccca	9000
atgatccgac	cagcaaaact	cgatgtactt	ccgaggaact	gatgtgcata	atgcatcagg	9060
ctggtacatt	agatccccgc	ttaccgcggg	caatatagca	acactaaaaa	ctcgatgtac	9120
ttccgaggaa	gcgcagtgca	taatgctgcg	cagtgttgcc	acataaccac	tatattaacc	9180
at'ttatctag	cggacgccaa	aaactcaatg	tatttctgag	gaagcgtggt	gcataatgcc	9240
acgcagcgtc	tgcataactt	ttattatttc	ttttattaat	caacaaaatt	ttgtttttaa	9300
catttcaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaag	ggaattccca	acttg'tttat	9360
tgcagcttat	aatgg'ttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	9420
ttttt'cactg	cattctagtt	gtgg'tttgtc	caaactcatc	aatgtatctt	atcatgtctg	9480
gatccg'tcga	gacgcgtg'cg	gccgcgggaa	ttcgatcttt	atgtgtaact	cttggctgaa	9540
g'tctttacac	caatgctggg	ggacatgtac	ctcccagggg	cccaggaaga	ctacgggagg	9600
ctacaccaac	gtcaatcaga	ggggcctgtg	tagctaccga	taagcggacc	ctcaagaggg	9660
cattagcaat	agtgtttata	aggccccctt	gttaacccta	aacgggtagc	atatgcttcc	9720
cgggtagtag	tatatactat	ccagactaac	cctaattcaa	tagcatatgt	tacccaacgg	9780
gaagcatatg	ctatcgaatt	agggttagta	aaagggtcct	aaggaacagc	gatatctccc	9840
accccatgag	ctgtcacggt	tttatttaca	tgggg'tcagg	attccacgag	ggtagtgaac	9900
catttttagtc	acaagggcag	tggctgaaga	tcaaggagcg	ggcagtgaac	tctcctgaat	9960
cttcgcctgc	ttcttcattc	tccttcg'ttt	agctaataga	ataactgctg	agttgtgaac	10020
agtaagg'tgt	atgtgaggtg	ctcgaaaaca	agg'tttcagg	tgacgcccc	agaataaaat	10080
ttggacgggg	ggttcag'tgg	tggcattgtg	ctatgacacc	aatataaccc	tcacaaaccc	10140
cttgggcaat	aaatactagt	gtaggaatga	aacattctga	atatct'ttaa	caatagaaat	10200
ccatgggg'tg	gggacaagcc	gtaaagactg	gatgtccatc	tcacacgaat	ttatggctat	10260
gggcaacaca	taatcctagt	gcaatatgat	actgggg'tta	ttaagatgtg	tcccaggcag	10320
ggaccaagac	aggtgaacca	tg'ttg'ttaca	ctctatt'tgt	aacaagggga	aagagagtgg	10380
acgccgacag	cagcggactc	cactgg'ttgt	ctctaacacc	cccga'aaatt	aaacggggct	10440

ccacgccaat ggggcccata aacaaagaca agtggccact ctttttttttg aaattgtgga 10500  
gtggggggcac gcgtcagccc ccacacgccg ccctgcggtt ttggactgta aaataagggt 10560  
gtaataactt ggctgattgt aaccccgcga accactgcgg tcaaaccact tgcccacaaa 10620  
accactaatg gcaccccggg gaatacctgc ataagtaggt gggcgggcca agataggggc 10680  
gcgattgctg cgatctggag gacaaattac acacacttgc gcctgagcgc caagcacagg 10740  
gttggtggtc ctcatattca cgaggctcgt gagagcacgg tgggctaatt ttgccatggg 10800  
tagcatatac tacccaaata tctggatagc atatgctatc ctaatctata tctgggtagc 10860  
ataggctatc ctaatctata tctgggtagc atatgctatc ctaatctata tctgggtagt 10920  
atatgctatc ctaatttata tctgggtagc ataggctatc ctaatctata tctgggtagc 10980  
atatgctatc ctaatctata tctgggtagt atatgctatc ctaatctgta tccgggtagc 11040  
atatgctatc ctaatagaga ttagggtagt atatgctatc ctaatttata tctgggtagc 11100  
atatactacc caaatatctg gatagcatat gctatcctaa tctatatctg ggtagcatat 11160  
gctatcctaa tctatatctg ggtagcatag gctatcctaa tctatatctg ggtagcatat 11220  
gctatcctaa tctatatctg ggtagtatat gctatcctaa tttatatctg ggtagcatag 11280  
gctatcctaa tctatatctg ggtagcatat gctatcctaa tctatatctg ggtagtatat 11340  
gctatcctaa tctgtatccg ggtagcatat gctatcctca tgcataata gtcagcatat 11400  
gatacccgat agtagagtgg gaggctatc ctttgcatat gccgccacct cccaaggggg 11460  
cgtgaatttt cgctgcttgt ctttttctg catgcatcac tagtgaattc gcggccgcgt 11520  
taggggtgtg aaagtcccca ggctccccag caggcagaag tatgcaaagc atgcatctca 11580  
attagtcagc aaccagggtg ggaaagtccc caggctcccc agcaggcaga agtatgcaaa 11640  
gcatgcatct caattagtca gcaaccatag tcccgccctt aactccgccc atcccgcctt 11700  
taactccgcc cagttccgcc cattctccgc cccatggctg actaattttt tttatttatg 11760  
cagaggccga ggccgcctcg gcctctgagc tattccagaa gtagtgagga ggcttttttg 11820  
gaggcctagg cttttgcaaa aagcttgcat gcctgcaggc cggccgccac gaccggtgcc 11880  
gccaccatcc cctgaccac gccctgacc cctcacaagg agacgacctt ccatgaccga 11940  
gtacaagccc acggtgcgcc tcgccacccg cgacgacgtc ccccgggccg tacgcacctt 12000  
cgccgcccgc ttcgccgact accccgccac gcgccacacc gtcgaccggg accgccacat 12060  
cgagcgggtc accgagctgc aagaactctt cctcacgcgc gtcgggctcg acatcggcaa 12120  
gggtgtgggtc gcggacgacg gcgccgcggt ggcggtctg accacgccgg agagcgctga 12180  
agcgggggag gtgttcgccg agatcggccc gcgcatggcc gagttgagcg gttcccggct 12240  
ggccgcgcag caacagatgg aaggcctcct ggccgcgcac cggcccaagg agcccgcgtg 12300  
gttcctggcc accgtcggcg tctcggccga ccaccagggc aagggtctg gcagcgccgt 12360  
cgtgctcccc ggagtgagg cgccgagcg cgccgggggtg cccgccttcc tggagacctc 12420  
cgcccccgc aacctccct tctacgagcg gctcggttc accgtcaccg ccgacgtcga 12480

ggtgcccga ggaaccgcgca cctggtgcat gacccgcaag cccggtgcct gacgcccgcc 12540  
 ccacgacccg cagcgcccga ccgaaaggag cgcacgaccc catggctccg accgaagcca 12600  
 cccggggcgg ccccgccgac cccgcacccg ccccgaggc ccaccgactc tagctagagg 12660  
 atcataatca gccataccac atttgttagag gttttacttg ctttaaaaaa cctccacac 12720  
 ctccccctga acctgaaaca taaaatgaat gcaattgttg ttgttaactt gtttattgca 12780  
 gcttataatg gttacaaata aagcaatagc atcacaatt tcacaaataa agcatttttt 12840  
 tctactgcatt ctagttgttg tttgtccaac gcgtcggtag cagcttttgt tccctttagt 12900  
 gagggttaat ttcgagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt 12960  
 atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctgggggtg 13020  
 cctaattgagt gagctaactc acattaattg cgtttcgctc actgcccgtt ttccagtcgg 13080  
 gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttg 13140  
 gtattggcg ctcttcgct tcctcgctca ctgactcgct gcgctcggtc gttcggctgc 13200  
 ggcgagcgg atcagctcac tcaaaggcgg taatacgggt atccacagaa tcaggggata 13260  
 acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaaccgt aaaaaggccg 13320  
 cgttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct 13380  
 caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa 13440  
 gctccctcgt gcgctctct gttccgacct tgccgcttac cggatacctg tccgcctttc 13500  
 tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcgggtg 13560  
 aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg 13620  
 ccttatccgg taactatcgt cttgagtcca acccggttag acacgactta tcgccactgg 13680  
 cagcagccac tggtaacagg attagcagag cgaggtagtg aggcgggtgct acagagttct 13740  
 tgaagtgggt gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc 13800  
 tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg 13860  
 ctggtagcgg tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc 13920  
 aagaagatcc tttgatcttt tctacggggg ctgacgctca gtggaacgaa aactcacgtt 13980  
 aagggatttt ggtcatgaga ttatcaaaaa ggtatctcac ctagatcctt ttaaattaaa 14040  
 aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat 14100  
 gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcattc atagttgcct 14160  
 gactccccgt cgtgtagata actacgatac gggagggcct accatctggc cccagtgtg 14220  
 caatgatacc gcgagacca cgctcaccgg ctccagattt atcagcaata aaccagccag 14280  
 ccggaagggc cgagcgaga agtggtcctg caactttatc cgctccatc cagtctatta 14340  
 attgttgccg ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttggtg 14400  
 ccattgctac aggcacgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg 14460  
 gttccaacg atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct 14520

ccttcggtcc tccgatcgtt gtcagaagta agttggccgc agtggtatca ctcatggta 14580  
 tggcagcact gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg 14640  
 gtgagtactc aaccaagtca ttctgagaat agtgatgacg gcgaccgagt tgctcttgcc 14700  
 cggcgctcaat acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg 14760  
 gaaaacgttc ttcggggcca aaactctcaa ggatcttacc gctgttgaga tccagttcga 14820  
 tgtaaccac tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg 14880  
 ggtgagcaaa aacaggaagg caaatgccc caaaaagg gataaggcg acacggaaat 14940  
 gttgaatact catactcttc ctttttcaat attattgaag catttatcag ggttattgtc 15000  
 tcatgagcgg atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca 15060  
 catttccccg aaaagtgcc c 15081

<210> 22  
 <211> 17753  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> cDNA

<400> 22  
 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtgggttac cgcagcgtga 60  
 ccgtacact tgccagcgcc ctacgcgccg ctcttttcgc tttcttcctt tcctttctcg 120  
 ccacgttcgc cggcggatcc ggccattagc catattattc attgggtata tagcataaat 180  
 caatattggc tattggccat tgcatacgtt gtatccatat cataatatgt acatttatat 240  
 tggctcatgt ccaacattac cgccatgttg acattgatta ttgactagtt attaatagta 300  
 atcaattacg gggtcattag ttcatagccc atatatggag ttccgcgtta cataacttac 360  
 ggtaaatggc ccgcctggct gaccgcccaa cgacccccgc ccattgacgt caataatgac 420  
 gtatgttccc atagtaacgc caataggac tttccattga cgtcaatggg tggagtattt 480  
 acggtaaact gccacttg cagtacatca agtgatcat atgccaagta cgccccctat 540  
 tgacgtcaat gacggtaaat ggccgcctg gcattatgcc cagtacatga cttatggga 600  
 ctttcctact tggcagtaca tctacgtatt agtcatcgct attaccatgg tgatgcgggt 660  
 ttggcagtac atcaatgggc gtggatagcg gtttgactca cggggatttc caagtctcca 720  
 cccattgac gtcaatggga gtttgtttt gcacaaaaat caacgggact ttcaaaaatg 780  
 tcgtaacaac tccgccccat tgacgcaa atggcggtagg cgtgtacggg gggagggtcta 840  
 tataagcaga gtcgttttag tgaaccgtat tgacggcgta gtacacacta ttgaatcaaa 900  
 cagccgacca attgcactac catcacaatg gagaagccag tagtaaacgt agacgtagac 960  
 cccagagtc cgtttgtcgt gcaactgcaa aaaagcttcc cgcaatttga ggtagtagca 1020  
 cagcaggtca ctccaaatga ccatgcta atgccagagcat tttcgcatct ggccagtaaa 1080  
 ctaatcgagc tggaggttcc taccacagcg acgatcttgg acataggcag cgcaccggct 1140

cgtagaatgt tttccgagca ccagtatcat tgtgtctgcc ccatgcgtag tccagaagac 1200  
 cgggaccgca tgatgaaata tgccagtaaa ctggcggaaa aagcgtgcaa gattacaaac 1260  
 aagaacttgc atgagaagat taaggatctc cggaccgtac ttgatacgcc ggatgctgaa 1320  
 acaccatcgc tctgctttca caacgatgtt acctgcaaca tgcgtgccga atattccgtc 1380  
 atgcaggacg tgtatatcaa cgctcccggga actatctatc atcaggctat gaaaggcgtg 1440  
 cggaccctgt actggattgg cttcgacacc acccagttca tgttctcggc tatggcaggt 1500  
 tcgtaccctg cgtacaacac caactgggcc gacgagaaag tccttgaagc gcgtaacatc 1560  
 ggactttgca gcacaaagct gagtgaaggt aggacaggaa aattgtcgat aatgaggaag 1620  
 aaggagttga agcccgggtc gcgggtttat ttctccgtag gatcgacact ttatccagaa 1680  
 cacagagcca gcttgagag ctggcatctt ccatcggtgt tccacttgaa tggaaagcag 1740  
 tcgtacactt gccgctgtga tacagtgggt agttgcgaag gctacgtagt gaagaaaatc 1800  
 accatcagtc ccgggatcac gggagaaacc gtgggatacg cggttacaca caatagcgag 1860  
 ggcttcttgc tatgcaaagt tactgacaca gtaaaaggag aacgggtatc gttccctgtg 1920  
 tgcacgtaca tcccggccac catatgcat gatagactg gtataatggc cacggatata 1980  
 tcacctgacg atgcacaaaa acttctgggt gggctcaacc agcgaattgt cattaacggt 2040  
 aggactaaca ggaacaccaa caccatgcaa aattaccttc tgccgatcat agcacaaggg 2100  
 ttcagcaaat gggctaagga gcgcaaggat gatcttgata acgagaaaat gctgggtact 2160  
 agagaacgca agcttacgta tggctgcttg tgggcgtttc gactaagaa agtacattcg 2220  
 ttttatcgcc cacctggaac gcagacctgc gtaaaagtcc cagcctcttt tagcgctttt 2280  
 cccatgtcgt ccgtatggac gacctctttg cccatgtcgc tgaggcagaa attgaaactg 2340  
 gcattgcaac caaagaagga ggaaaaactg ctgcaggtct cggaggaatt agtcatggag 2400  
 gccaaaggct cttttgagga tgctcaggag gaagccagag cggagaagct ccgagaagca 2460  
 cttccaccat tagtggcaga caaaggcatc gaggcagccg cagaagttgt ctgcgaagtg 2520  
 gaggggctcc aggcggacat cggagcagca ttagttgaaa ccccgcgagg tcacgtaagg 2580  
 ataatacctc aagcaaatga ccgtatgatc ggacagtata tcgttgtctc gccaaactct 2640  
 gtgctgaaga atgcaaaact cgcaccagcg caccgctag cagatcaggt taagatcata 2700  
 acacactccg gaagatcagg aaggtagcg gtcgaaccat acgacgctaa agtactgatg 2760  
 ccagcaggag gtgccgtacc atggccagaa ttcctagcac tgagtgaag cgccacgtta 2820  
 gtgtacaacg aaagagagtt tgtgaaccgc aaactatacc acattgccat gcatggcccc 2880  
 gccagaata cagaagagga gcagtacaag gttacaaagg cagagcttgc agaaacagag 2940  
 tacgtgtttg acgtggacaa gaagcgttgc gtttaagaagg aagaagcctc aggtctgggtc 3000  
 ctctcgggag aactgaccaa cctccctat catgagctag ctctggaggg actgaagacc 3060  
 cgacctgcgg tcccgtacaa ggtcgaaaca ataggagtga taggcacacc ggggtcgggc 3120  
 aagtcagcta ttatcaagtc aactgtcacg gcacgagatc ttgttaccag cggaagaaa 3180

gaaaattgtc gcgaaattga ggccgacgtg ctaagactga ggggtatgca gattacgtcg	3240
aagacagtag attcggttat gctcaacgga tgccacaaag ccgtagaagt gctgtacgtt	3300
gacgaagcgt tcgctgcca cgcaggagca ctacttgcct tgattgctat cgtcaggccc	3360
cgcaagaagg tagtactatg cggagacccc atgcaatgcg gattcttcaa catgatgcaa	3420
ctaaaggtag atttcaatca ccctgaaaaa gacatatgca ccaagacatt ctacaagtat	3480
atctcccggc gttgcacaca gccagttaca gctattgtat cgacactgca ttacgatgga	3540
aagatgaaaa ccacgaaccc gtgcaagaag aacattgaaa tcgatattac aggggccaca	3600
aagccgaagc caggggatat catcctgaca tgtttccgcg ggtgggttaa gcaattgcaa	3660
atcgactatc ccggacatga agtaatgaca gccgcggcct cacaagggct aaccagaaaa	3720
ggagtgtatg ccgtccggca aaaagtcaat gaaaacccac tgtacgcgat cacatcagag	3780
catgtgaacg tgttgctcac ccgactgag gacaggctag tgtggaaaac cttgcagggc	3840
gacccatgga ttaagcagct cactaacata cctaaaggaa actttcaggc tactatagag	3900
gactgggaag ctgaacacaa gggaataatt gctgcaataa acagccccac tcccgtgcc	3960
aatccgttca gctgcaagac caacgtttgc tgggcgaaag cattggaacc gatactagcc	4020
acggccggtg tcgtacttac cggttgccag tggagcgaac tgttcccaca gtttgcggt	4080
gacaaaccac attcggccat ttacgcctta gacgtaattt gcattaagtt tttcggcatg	4140
gacttgacaa gcggactgtt ttctaaacag agcatccac taacgtacca tcccgcgat	4200
tcagcgaggc cggtagctca ttgggacaac agcccaggaa cccgcaagta tgggtacgat	4260
cacgccattg ccgccgaact ctcccgtaga tttccggtgt tccagctagc tgggaagggc	4320
acacaacttg atttgagac ggggagaacc agagttatct ctgcacagca taacctggtc	4380
ccggtgaacc gcaatcttcc tcacgcctta gtccccgagt acaaggagaa gcaacccggc	4440
ccggtcgaaa aattcttgaa ccagttcaaa caccactcag tacttgtggt atcagaggaa	4500
aaaattgaag ctccccgtaa gagaatcgaa tggatcgccc cgattggcat agccggtgca	4560
gataagaact acaacctggc tttcgggttt ccgccgagg cacggtacga cctggtgttc	4620
atcaacattg gaactaaata cagaaaccac cactttcagc agtgcggaaga ccatgcggcg	4680
accttaaaaa ccctttcgcg ttcggccctg aattgtttta actcaggagg caccctcgtg	4740
gtgaagtcct atggctacgc cgaccgcaac agtgaggacg tagtcaccgc tcttgccaga	4800
aagtttgtca ggggtgtctg agcgagacca gattgtgtct caagcaatac agaaatgtac	4860
ctgattttcc gacaactaga caacagccgt acacggcaat tcccccgca ccatctgaat	4920
tgcgtgattt cgtccgtgta tgagggtaca agagatggag ttggagccgc gccgtcatac	4980
cgcacaaaaa gggagaatat tgctgactgt caagaggaag cagttgtcaa cgcagccaat	5040
ccgctgggta gaccaggcga aggagtctgc cgtgccatct ataaacgttg gccgaccagt	5100
tttaccgatt cagccacgga gacaggcacc gcaagaatga ctgtgtgcct aggaaagaaa	5160
gtgatccacg cggtcggccc tgatttccgg aagcaccag aagcagaagc cttgaaattg	5220



ctacaaaacg cctaccatgc agtggcagac ttagtaaagc aacataacat caagtctgtc	5280
gccattccac tgctatctac aggcatttac gcagccggaa aagaccgcct tgaagtatca	5340
cttaactgct tgacaaccgc gctagacaga actgacgcgg acgtaaccat ctattgcctg	5400
gataagaagt ggaaggaaag aatcgacgcg gcaactcaac ttaaggagtc tgtaacagag	5460
ctgaaggatg aagatatgga gatcgacgat gagttagtat ggattcatcc agacagttgc	5520
ttgaaggga gaaagggatt cagtactaca aaaggaaaat tgtattcgta cttcgaaggc	5580
accaaattcc atcaagcagc aaaagacatg gcggagataa aggtcctgtt ccctaattgac	5640
caggaaagta atgaacaact gtgtgcctac atattgggtg agaccatgga agcaatccgc	5700
gaaaagtgcc cggtcgacca taaccgcgcg tctagccgcg ccaaacggt gccgtgcctt	5760
tgcatgtatg ccatgacgcc agaaagggtc cacagactta gaagcaataa cgtcaaagaa	5820
gttacagtat gtcctccac ccccttctc aagcacaaaa ttaagaatgt tcagaagggt	5880
cagtgcacga aagtagtctt gtttaatccg cacactcccg cattcgttcc cgcccgtaa	5940
tacatagaag tgccagaaca gcctaccgct cctcctgcac aggccgagga ggccccgaa	6000
gttgtagcga caccgtcacc atctacagct gataacacct cgcttgatgt cacagacatc	6060
tactggata tggatgacag tagcgaaggc tcactttttt cgagcttttag cggatcggac	6120
aactctatta ctagtatgga cagttgggtc tcaggacctt gttcactaga gatagtagac	6180
cgaaggcagg tgggtgggtg tgacgttcat gccgtccaag agcctgcccc tattccaccg	6240
ccaaggctaa agaagatggc ccgcctggca gcggcaagaa aagagcccac tccaccggca	6300
agcaatagct ctgagtcctt ccacctctt tttgggtggg tatccatgtc cctcggatca	6360
atcttcgacg gagagacggc ccgcctggca gcggtacaac ccctggcaac aggccccacg	6420
gatgtgccta tgtctttcgg atcgttttcc gacggagaga ttgatgagct gagccgcaga	6480
gtaactgagt ccgaacccgt cctgtttgga tcatttgaac cgggcgaagt gaactcaatt	6540
atatcgtccc gatcagccgt atcttttctt ctacgcaagc agagacgtag acgcaggagc	6600
aggaggactg aatactgact aaccggggta ggtgggtaca tattttcgac ggacacaggc	6660
cctgggcact tgcaaaagaa gtccgttctg cagaaccagc ttacagaacc gaccttgag	6720
cgcaatgtcc tggaaagaat tcatgccccg gtgctcgaca cgtcgaaaga ggaacaactc	6780
aaactcaggt accagatgat gccaccgaa gccaaacaaa gtaggtacca gtctcgtaaa	6840
gtagaaaatc agaaagccat aaccactgag cgactactgt caggactacg actgtataac	6900
tctgccacag atcagccaga atgctataag atcacctatc cgaaaccatt gtactccagt	6960
agcgtaccgg cgaactactc cgatccacag ttcgctgtag ctgtctgtaa caactatctg	7020
catgagaact atccgacagt agcatcttat cagattactg acgagtacga tgcttacttg	7080
gatatggtag acgagacagt cgcgtgcctg gatactgcaa cttctgccc cgctaagctt	7140
agaagttacc cgaaaaaaca tgagtataga gccccgaata tccgcagtgc ggttccatca	7200
gcgatgcaga acacgctaca aaatgtgctc attgccgcaa ctaaaagaaa ttgcaacgctc	7260

acgcagatgc	gtgaactgcc	aacactggac	tcagcgacat	tcaatgtcga	atgcttttcga	7320
aaatatgcat	gtaatgacga	gtattgggag	gagttcgctc	ggaagccaat	taggattacc	7380
actgagtttg	tcaccgcata	tgtagctaga	ctgaaaggcc	ctaaggccgc	cgcactat	7440
gcaaagacgt	ataatttggg	cccattgcaa	gaagtgccta	tggatagatt	cgtcatggac	7500
atgaaaagag	acgtgaaagt	tacaccaggc	acgaaacaca	cagaagaaaag	accgaaagta	7560
caagtgatac	aagccgcaga	acccttggcg	actgcttact	tatgcgggat	tcaccgggaa	7620
ttagtgcgta	ggcttacggc	cgtcttgctt	caaacattc	acacgctttt	tgacatgtcg	7680
gcggaggatt	ttgatgcaat	catagcagaa	cacttcaagc	aaggcgaccc	ggtactggag	7740
acggatatcg	catcattcga	caaaagccaa	gacgacgcta	tggcggttaac	cggctctgatg	7800
atcttggagg	acctgggtgt	ggatcaacca	ctactcgact	tgatcgagtg	cgcctttgga	7860
gaaatatcat	ccacccatct	acctacgggt	actcgtttta	aattcggggc	gatgatgaaa	7920
tccggaatgt	tcctcacact	ttttgtcaac	acagttttga	atgtcgttat	cgccagcaga	7980
gtactagaag	agcggcttaa	aacgtccaga	tgtgcagcgt	tcattggcga	cgacaacatc	8040
atacatggag	tagtatctga	caaagaaatg	gctgagagggt	gcgccacctg	gctcaacatg	8100
gaggttaaga	tcatcgacgc	agtcacgggt	gagagaccac	cttacttctg	cggcggattt	8160
atcttgcaag	attcgggttac	ttccacagcg	tgccgcgtgg	cggatcccct	gaaaaggctg	8220
tttaagttgg	gtaaaccgct	cccagccgac	gacgagcaag	acgaagacag	aagacgcgct	8280
ctgctagatg	aaacaaaggc	gtgggtttaga	gtaggtataa	caggcacttt	agcagtggcc	8340
gtgacgaccc	ggtatgagggt	agacaatatt	acacctgtcc	tactggcatt	gagaactttt	8400
gcccagagca	aaagagcatt	ccaagccatc	agaggggaaa	taaagcatct	ctacgggtggt	8460
cctaaatagt	cagcatagta	catttcatct	gactaatacc	acaacaccac	cacctctaga	8520
ttaattaact	cgaggcgcgc	cgggccccaga	ggatctttgt	gaaggaacct	tacttctgtg	8580
gtgtgacata	attggacaaa	ctacctacag	agatttaaag	ctctaaggta	aatataaaat	8640
ttttaagtgt	ataatgtgtt	aaactactga	ttctaattgt	ttgtgtat	tagattccaa	8700
cctatggaac	tgatgaatgg	gagcagtgggt	ggaatgcctt	taatgaggaa	aacctgtttt	8760
gctcagaaga	aatgccatct	agtgatgatg	aggctactgc	tgactctcaa	cattctactc	8820
ctccaaaaaa	gaagagaaaag	gtagaagacc	ccaaggactt	tccttcagaa	ttgctaagtt	8880
ttttgagtca	tgctgtgttt	agtaatagaa	ctcttgcttg	ctttgctatt	tacaccacaa	8940
aggaaaaagc	tgacttgcta	tacaagaaaa	ttatggaaaa	atatttgatg	tagcggccca	9000
atgatccgac	cagcaaaact	cgatgtactt	ccgaggaact	gatgtgcata	atgcatcagg	9060
ctggtacatt	agatccccgc	ttaccgcggg	caatatagca	acactaaaaa	ctcgatgtac	9120
ttccgaggaa	gcgcagtgca	taatgtctgcg	cagtgttgcc	acataaccac	tatattaacc	9180
atztatctag	cggacgccaa	aaactcaatg	tatttctgag	gaagcgtgggt	gcataatgcc	9240
acgcagcgtc	tgcataactt	ttattatttc	ttttattaat	caacaaaatt	ttgtttttaa	9300

catttcaaaa	aaaaaaaaa	aaaaaaaaa	aaaaaaaagg	ggaattccca	acttgtttat	9360
tgcagcttat	aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	9420
tttttactg	cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	atcatgtctg	9480
gatccgtcga	gacgcgtgcg	gccgcgggaa	ttcgaatggt	ctttcatcct	ggagcagact	9540
ttgcagtctg	tggactgcaa	cacaacattg	cctttatgtg	taactcttgg	ctgaagctct	9600
tacaccaatg	ctgggggaca	tgtacctccc	agggggccag	gaagactacg	ggagggtaca	9660
ccaacgtcaa	tcagaggggc	ctgtgtagct	accgataagc	ggaccctcaa	gagggcatta	9720
gcaatagtgt	ttataaggcc	ccttggttaa	ccctaaacgg	gtagcatatg	cttccccgggt	9780
agtagtatat	actatccaga	ctaaccctaa	ttcaatagca	tatgttacct	aacgggaagc	9840
atatgctatc	gaattagggg	tagtaaaagg	gtcctaagga	acagcgatat	ctcccacccc	9900
atgagctgtc	acggttttat	ttacatgggg	tcaggattcc	acgagggtag	tgaaccattt	9960
tagtcacaag	ggcagtggtc	gaagatcaag	gagcgggcag	tgaactctcc	tgaatcttcg	10020
cctgcttctt	cattctcctt	cgttttagcta	atagaataac	tgctgagttg	tgaacagtaa	10080
ggtgtatgtg	aggtgctcga	aaacaagggt	tcagggtgacg	cccccagaat	aaaatttgga	10140
cgggggggtc	agtgggtggc	ttgtgctatg	acaccaatat	aaccctcaca	aacccttggg	10200
gcaataaata	ctagtgtagg	aatgaaacat	tctgaatatc	tttaacaata	gaaatccatg	10260
gggtggggac	aagccgtaaa	gactggatgt	ccatctcaca	cgaatttatg	gctatgggca	10320
acacataatc	ctagtgcaat	atgatactgg	ggttattaag	atgtgtccca	ggcagggacc	10380
aagacaggtg	aaccatgttg	ttacactcta	tttghtaaca	ggggaaagag	agtggacgcc	10440
gacagcagcg	gactccactg	gttgctctcta	acacccccga	aaattaaacg	gggctccacg	10500
ccaatggggc	ccataaaca	agacaagtgg	ccactctttt	ttttgaaatt	gtggagtggg	10560
ggcacgcgtc	agccccca	cgccgccctg	cggttttgga	ctgtaaaata	aggggtgtaat	10620
aacttggtcg	attgtaaccc	cgctaaccac	tgcggtcaaa	ccacttgccc	acaaaaccac	10680
taatggcacc	ccgggggaata	cctgcataag	taggtgggcg	ggccaagata	ggggcgcgat	10740
tgctgcgatc	tggaggacaa	attacacaca	cttgcgccctg	agcgccaagc	acaggggttg	10800
tggtcctcat	attcacgagg	tcgctgagag	cacgggtgggc	taatgttgcc	atgggtagca	10860
tatactaccc	aaatatctgg	atagcatatg	ctatccta	ctatatctgg	gtagcatagg	10920
ctatccta	ctatatctgg	gtagcatatg	ctatccta	ctatatctgg	gtagtatatg	10980
ctatccta	ttatatctgg	gtagcatagg	ctatccta	ctatatctgg	gtagcatatg	11040
ctatccta	ctatatctgg	gtagtatatg	ctatccta	ctgtatccgg	gtagcatatg	11100
ctatccta	agagattagg	gtagtatatg	ctatccta	ttatatctgg	gtagcatata	11160
ctacccaaat	atctggatag	catatgctat	cctaattctat	atctgggtag	catatgctat	11220
cctaattctat	atctgggtag	cataggctat	cctaattctat	atctgggtag	catatgctat	11280
cctaattctat	atctgggtag	tatatgctat	cctaatttat	atctgggtag	cataggctat	11340

cctaattctat atctgggtag catatgctat cctaattctat atctgggtag tatatgctat 11400  
 cctaattctgt atccgggtag catatgctat cctcatgcat atacagtcag catatgatac 11460  
 ccagtagtag agtgggagtg ctatcctttg catatgccgc cacctcccaa gggggcgtag 11520  
 attttcgctg cttgtccttt tcctgcatgc tggttgctcc cattcttagg tgaatttaag 11580  
 gaggccaggc taaagccgtc gcatgtctga ttgctcacca ggtaaagtgc gctaattgtt 11640  
 tccaacgcga gaaggtgttg agcgcgagc tgagtgcgt gacaacatgg gtatgcccaa 11700  
 ttgccccatg ttgggaggac gaaaatggtg acaagacaga tggccagaaa tacaccaaca 11760  
 gcacgcatga tgtctactgg ggatttattc tttagtgcgg gggaatacac ggcttttaat 11820  
 acgattgagg gcgtctccta acaagttaca tcactcctgc ccttcctcac cctcatctcc 11880  
 atcacctcct tcctctccgt catctccgtc atcacctcc gcggcagccc cttccaccat 11940  
 aggtggaaac caggaggga aatctactcc atcgtaaaag ctgcacacag tcaccctgat 12000  
 attgcaggta ggagcgggt ttgtcataac aaggtcctta atcgcatcct tcaaacctc 12060  
 agcaaatata tgagtttga aaaagaccat gaaataacag acaatggact cccttagcgg 12120  
 gccaggttgt gggccgggtc caggggccat tccaaagggg agacgactca atggtgtaag 12180  
 acgacattgt ggaatagcaa gggcagttcc tcgccttagg ttgtaaaggg aggtcttact 12240  
 acctccatat acgaacacac cggcgaccca agttccttcg tcggtagtcc tttctacgtg 12300  
 actcctagcc aggagagctc ttaaaccctc tgcaatgttc tcaaatttcg gggttgaacc 12360  
 tccttgacca cgatgctttc caaaccaccc tccttttttg cgctgcctc catcacctg 12420  
 accccgggt ccagtgttg ggccttctcc tgggtcatct gcggggccct gctctatcgc 12480  
 tcccgggggc acgtcaggct caccatctgg gccaccttct tgggtgtatt caaaataatc 12540  
 ggcttcccct acagggtgga aaaatggcct tctacctgga gggggcctgc gcggtggaga 12600  
 cccggatgat gatgactgac tactgggact cctgggcctc ttttctccac gtccacgacc 12660  
 tctccccctg gctctttcac gacttcccc cctggctctt tcacgtctc taccctggcg 12720  
 gcctccacta cctcctcgac cccggcctc actacctcct cgaccccggc ctccactgcc 12780  
 tcctcgaccc cggcctccac ctctgtctc tgccctcct gctcctgcc ctctcctgc 12840  
 tcctgcccc cctgcccc cgtctcctgc cctcctgcc cctcctgctc ctgccccctc 12900  
 tgccccctc gctcctgcc ctctgcccc tcctcctgct cctgcccc cctgccccctc 12960  
 tcctgctcct gccccctc cctcctgc tcctgcccc cctgcccc cctgctcctgc 13020  
 cctcctgcc cctcctgctc ctgccccctc tgctcctgcc cctcctgctc ctgccccctc 13080  
 tgctcctgcc cctcctgccc ctctgcccc tcctcctgct cctgcccc cctgctcctgc 13140  
 cctcctgcc cctcctgccc ctctgctcc tgccccctc cctgctcctg cccctcctgc 13200  
 cctcctgcc cctcctcctg ctctgcccc tcctgcccc cctcctgctc ctgccccctc 13260  
 tcctgctcct gccccctc cccctcctgc cctcctcct gctcctgcc ctctgcccc 13320  
 tcctcctgct cctgcccc ctctgctcc tgccccctc gccccctc cccctcctc 13380

tgctcctgcc	cctcctcctg	ctcctgcccc	tcctgcccct	cctgcccctc	ctgcccctcc	13440
tcctgctcct	gccccctctc	ctgctcctgc	ccctcctgct	cctgcccctc	ccgctcctgc	13500
tcctgctcct	gttccaccgt	gggtcccttt	gcagccaatg	caacttggac	gtttttgggg	13560
tctccggaca	ccatctctat	gtcttgcccc	tgatcctgag	ccgcccgggg	ctcctggtct	13620
tccgcctcct	cgctcctgct	ctcttccccg	tcctcgtcca	tggttatcac	cccctcttct	13680
ttgaggcca	ctgccgccgg	agccttcttg	tccagatgtg	tctcccttct	ctcctaggcc	13740
atttccaggt	cctgtacctg	gccccctgct	agacatgatt	cacactaaaa	gagatcaata	13800
gacatcttta	ttagacgacg	ctcagtgaat	acagggagtg	cagactcctg	ccccctccaa	13860
cagccccccc	accctcatcc	ccttcatggt	cgctgtcaga	cagatccagg	tctgaaaatt	13920
ccccatcctc	cgaaccatcc	tcgtcctcat	caccaattac	tcgcagcccc	gaaaactccc	13980
gctgaacatc	ctcaagattt	gcgtcctgag	cctcaagcca	ggcctcaa	tcctcgtccc	14040
cctttttgct	ggacggtagg	gatggggatt	ctcgggaccc	ctcctcttcc	tcttcaagggt	14100
caccagacag	agatgctact	ggggcaacgg	aagaaaagct	gggtgcggcc	tgtgaggatc	14160
agcttatatc	actagtgaat	tcgcggccgc	gttaggggtg	ggaaagtccc	caggctcccc	14220
agcaggcaga	agtatgcaaa	gcatgcatct	caattagtca	gcaaccaggt	gtggaaagtc	14280
cccaggctcc	ccagcaggca	gaagtatgca	aagcatgcat	ctcaattagt	cagcaaccat	14340
agtcccgc	ctaactccgc	ccatcccgc	cctaactccg	cccagttccg	cccattctcc	14400
gccccatggc	tgactaattt	tttttattta	tgcagaggcc	gaggccgcct	cggcctctga	14460
gctattccag	aagtagtgag	gaggcttttt	tggaggccta	ggcttttgca	aaaagcttgc	14520
atgcctgcag	gtcggccgcc	acgaccggtg	ccgccaccat	cccctgaccc	acgcccctga	14580
cccctcacia	ggagacgacc	ttccatgacc	gagtacaagc	ccacgggtgcg	cctcgccacc	14640
cgcgacgacg	tcctccgggc	cgtacgcacc	ctcgccgccg	cggtcgccga	ctaccccgcc	14700
acgcgccaca	ccgtcgaccc	ggaccgccac	atcgagcggg	tcaccgagct	gcaagaactc	14760
ttcttcacgc	gcgtcgggct	cgacatcggc	aagggtgtgg	tcgcggacga	cggcgccgcg	14820
gtggcggtct	ggaccacgcc	ggagagcgtc	gaagcggggg	cggtgttcgc	cgagatcggc	14880
ccgcgcagtg	ccgagttgag	cggttcccgg	ctggccgcgc	agcaacagat	ggaaggcctc	14940
ctggcgccgc	accggcccaa	ggagcccgcg	tggttcctgg	ccaccgtcgg	cgtctcgccc	15000
gaccaccagg	gcaagggtct	gggcagcgcc	gtcgtgctcc	ccggagtggg	ggcgcccgag	15060
cgcgccgggg	tgcccgctt	cctggagacc	tccgcgcccc	gcaacctccc	cttctacgag	15120
cggctcggct	tcaccgtcac	cgccgacgtc	gagggtgccc	aaggaccgcg	cacctggtgc	15180
atgaccgcga	agcccgggtg	ctgacgcccc	ccccacgacc	cgacgcgccc	gaccgaaagg	15240
agcgacgac	ccatgggtc	cgaccgaagc	caccgggggc	ggccccgccc	accccgaccc	15300
cgcccccgag	gcccaccgac	tctagctaga	ggatcataat	cagccatacc	acattttag	15360
aggttttact	tgctttaaaa	aacctccac	acctccccct	gaacctgaaa	cataaaatga	15420

32

aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	actcgtgcac	ccaactgatc	17520
ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	aaaacaggaa	ggcaaaatgc	17580
cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	ctcatactct	tcctttttca	17640
atattattga	agcattttatc	aggggttattg	tctcatgagc	ggatacatat	ttgaatgtat	17700
ttagaaaaat	aaacaaatag	gggttcgcgcg	cacattttccc	cgaaaagtgc	cac	17753

<b>&lt;210&gt;</b>	<b>23</b>
<b>&lt;211&gt;</b>	<b>33</b>
<b>&lt;212&gt;</b>	<b>DNA</b>
<b>&lt;213&gt;</b>	<b>Artificial Sequence</b>

**<220>**  
**<223> primer**

**<400> 23**  
gtcatgaagc ttgccaccgc tcagagcacc ttc **33**

<210>	24
<211>	31
<212>	DNA
<213>	Artificial Sequence

**<220>**  
**<223> primer**

**<400> 24**  
**gacttctaga ctggggcggg gtgggatgaa c** **31**

<b>&lt;210&gt;</b>	<b>25</b>
<b>&lt;211&gt;</b>	<b>33</b>
<b>&lt;212&gt;</b>	<b>DNA</b>
<b>&lt;213&gt;</b>	<b>Artificial sequence</b>

**<220>**  
**<223> primer**

**<400> 25**  
**gtcatgaagc ttctcggcga ccagaacacc ttc** **33**

<b>&lt;210&gt;</b>	<b>26</b>
<b>&lt;211&gt;</b>	<b>31</b>
<b>&lt;212&gt;</b>	<b>DNA</b>
<b>&lt;213&gt;</b>	<b>Artificial sequence</b>

<220>  
<223> primer

**<400> 26**  
**gacttctaga ctggggcggg gtgggatgaa c** **31**